



UNIVERSIDADE NOVA DE LISBOA
INSTITUTO DE HIGIENE E MEDICINA TROPICAL



MESTRADO EM CIÊNCIAS BIOMÉDICAS

(especialidade de Biologia Molecular em Medicina Tropical e Internacional)

CARACTERIZAÇÃO GENÉTICA DE ESTIRPES DO VÍRUS GBV-C
CIRCULANTES NA REGIÃO
METROPOLITANA DE LISBOA

CRISTINA VIEIRA BRANCO

2010



UNIVERSIDADE NOVA DE LISBOA
INSTITUTO DE HIGIENE E MEDICINA TROPICAL



CARACTERIZAÇÃO GENÉTICA DE ESTIRPES DO VÍRUS GBV-C
CIRCULANTES NA REGIÃO
METROPOLITANA DE LISBOA

CRISTINA VIEIRA BRANCO

*Tese apresentada para a obtenção do grau de Mestre em Ciências Biomédicas
(especialidade de Biologia Molecular em Medicina Tropical e Internacional)*

Orientador:

Professor Doutor Ricardo Parreira

2010

Apresentação em Congresso, (Formato *Poster*): (ver Anexo VI)

BRANCO, C., ESTEVES, A., PIEDADE, J. & PARREIRA, R., 2009. Prevalence of infection and genetic analysis of GB virus C in a population group comprising a large number of intravenous drug users (IDUs).

Book of Abstracts of Microbiotec09, Poster Session: S3-Clinical Microbiology and Epidemiology, 28-30 Novembro, 2009, Vilamoura, Portugal.

Microbiotec09, 2009, Centro de Congressos do Tivoli Marina Hotel, Vilamoura, Portugal

Publicação no âmbito deste trabalho: (ver Anexo VII)

BRANCO, C., ESTEVES, A., PIEDADE, J. & PARREIRA, R., 2010. A new genotype 2 subcluster identified among GBV-C strains circulating in the Lisbon metropolitan area of Portugal. *Journal of Medical Virology*, **82**(3):452-9.

AGRADECIMENTOS

Uma dissertação, apesar de ser um trabalho individual, devido ao seu objectivo académico, envolve uma interligação e cooperação com diversas pessoas, que não poderia deixar de ser mencionada, pelo que expresso aqui os meus sinceros agradecimentos a todos aqueles que, tornaram a realização deste trabalho possível.

Um especial agradecimento ao Professor Doutor Ricardo Parreira, orientador desta tese, sempre presente nos momentos cruciais, pelo apoio, paciência, dedicação, pelas críticas e sugestões relevantes feitas durante a orientação, e pela disponibilidade demonstrada em todas as fases que levaram à concretização deste trabalho.

Gostaria ainda de agradecer,

À Professora Doutora Aida Esteves, pela atenção, estímulo e ensinamentos transmitidos.

Ao Professor Doutor João Piedade, pelos conselhos e motivação.

À Raquel, colega de mestrado e acima de tudo, uma grande amiga, presente do início ao fim destes dois longos anos de percurso académico.

À Sandra Castro e Susana Valente, pela amizade, companheirismo, partilha de conhecimentos e momentos de convívio.

À Teresa Venneno, pela experiência, preciosa ajuda laboratorial, e pelo sorriso com que sempre me recebeu.

Às técnicas da Unidade de Virologia, IHMT, Fernanda e Cidália, pelo empenhamento e prontidão com que acolheram todos os pedidos “urgentes”.

Às colegas da Unidade de Micobactérias, IHMT, pelos alegres momentos de descontração e camaradagem.

É com muita satisfação que expresso aqui o mais profundo agradecimento,

À minha família, pelo carinho, apoio incondicional e acima de tudo por acreditarem em mim.

Ao Ferdi, pelo que construímos juntos neste último ano,

Dedico esta tese à minha mãe.

ÍNDICE GERAL

AGRADECIMENTOS.....	I
ÍNDICE GERAL.....	III
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABELAS.....	X
ABREVIATURAS.....	XII
RESUMO.....	XVI
ABSTRACT.....	XVII

1. INTRODUÇÃO

1.1. Nota introdutória.....	2
1.2. Os vírus GB: uma perspectiva histórica.....	4
1.3. O vírus GBV-C.....	6
1.3.1. Características gerais da partícula viral e estrutura genómica	6
1.3.2. Tropismo e replicação do vírus GBV-C.....	9
1.3.3. Ciclo replicativo.....	10
1.3.4. Infecção e possível associação a doença.....	11
1.4. Impacto do GBV-C nas co-infecções com HIV e/ou HCV.....	12
1.4.1. Relevância biológica.....	12
1.4.2. Dinâmica da co-infecção GBV-C/HIV.....	14
1.4.2.1. Mecanismos propostos para a interacção GBV-C/HIV.....	16
1.4.3. Co-infecção com HCV.....	19
1.5. Caracterização epidemiológica do vírus GBV-C.....	21
1.5.1. Distribuição geográfica e variabilidade genotípica.....	21
1.5.1.1. Co-evolução dos vírus GBV-C com os seus hospedeiros humanos.....	23
1.5.2. Vias de transmissão.....	24
1.5.3. Taxas de infecção pelo vírus GBV-C em diferentes grupos populacionais.....	25
1.5.4. Marcadores moleculares e detecção de GBV-C.....	28
1.6. Notas finais/Objectivos.....	29

2. MATERIAL E MÉTODOS

2.1.	Caracterização da população estudada.....	32
2.1.1.	Produto biológico: colheita e preparação da amostra.....	33
2.2.	Extracção de RNA total a partir de plasma humano.....	34
2.3.	Síntese, por transcrição reversa (RT), de DNA complementar (cDNA) a partir de RNA total.....	35
2.4.	Delineamento de oligonucleótidos sintéticos (<i>primers</i>) iniciadores de PCR.....	36
2.5.	Amplificação de sequências genómicas do vírus GBV-C por <i>nested</i> -PCR.....	38
2.5.1.	Detecção da presença do vírus GBV-C: amplificação da região conservada 5'UTR.....	38
2.5.2.	Amplificação da região 5'UTR/E1 do genoma do GBV-C.....	40
2.5.3.	Amplificação das regiões E1/E2 e NS5A/NS5B do genoma do GBV-C.....	41
2.6.	Análise de moléculas de DNA por electroforese em gel de agarose.....	42
2.7.	Purificação dos produtos de PCR.....	42
2.8.	Clonagem molecular de fragmentos do genoma de GBV-C.....	44
2.8.1.	Preparação de células competentes de <i>E.coli</i>	44
2.8.2.	Clonagem molecular nos vectores PGEM®-T e PGEMT®-Easy.....	44
2.8.3.	Transformação de <i>E.coli</i> por choque térmico.....	45
2.8.4.	Extracção do DNA plasmídico por lise alcalina.....	46
2.8.5.	Digestão enzimática de DNA.....	47
2.8.6.	Análise de moléculas recombinantes de DNA através da amplificação, por PCR, dos fragmentos clonados.....	47
2.8.7.	Extracção de DNA plasmídico em média escala.....	48
2.9.	Ensaio de PCR em Tempo Real (<i>Real Time</i> PCR).....	48
2.9.1	Construção de uma recta padrão de amplificação.....	50
2.9.2.	Ensaio de PCR em Tempo Real com amostras biológicas.....	51
2.9.2.1.	Ensaio de PCR em Tempo Real com <i>SYBR Green I</i>	52
2.9.2.2.	Ensaio de PCR em Tempo Real com sonda <i>TaqMan</i>	52
2.9.2.3.	Ensaio de heminested-PCR com amostras testadas por PCR em	

Tempo Real com sonda <i>TaqMan</i>	52
2.10. Sequenciação de DNA.....	53
2.11. Análise bioinformática.....	53
2.11.1. Análise de sequências nucleotídicas/proteicas.....	53
2.11.2. Análise filogenética com base em alinhamentos de sequências nucleotídicas	54
2.11.3. Análise estatística.....	56
2.11.4. Previsão de estruturas secundárias em moléculas de RNA.....	57
 3. RESULTADOS	
 3.1. Detecção de genomas do vírus GBV-C em plasma humano (virémia).....	59
3.2. Amplificação das regiões codificantes 5'UTR/E1, E1/E2, NS5A/NS5B.....	62
3.2.1. Análise da região 5'UTR.....	62
3.2.2.1. Análise da sequência da região 5'UTR/E1 do genoma de GBV-C: possível impacto da formação de estruturas secundárias.....	63
3.2.2. Análise das regiões E1/E2 e NS5A/NS5B.....	68
3.3. Clonagem molecular de sequências genômicas de GBV-C em vetores plasmídicos	70
3.4. Análise de sequências nucleotídicas.....	73
3.4.1. Reconstrução de filogenias baseadas na análise de sequências de GBV-C.....	73
3.4.2. Relações filogenéticas entre estirpes de GBV-C estimadas pela análise da região E1/E2.....	74
3.4.2.1. Caracterização genética do agrupamento G2*.....	81
3.4.3. Relações filogenéticas entre estirpes de GBV-C estimadas pela análise da região NS5A/NS5B Análise das regiões E1/E2 e NS5A/NS5B.....	86
3.5. Análise semi-quantitativa da virémia de GBV-C por PCR em Tempo Real.....	98
3.5.1. Construção de uma recta-padrão.....	98
3.5.2. Quantificação relativa do genoma de GBV-C nas amostras biológicas estudadas.....	100

4. DISCUSSÃO DE RESULTADOS.....	103
Determinação da prevalência de GBV-C na região metropolitana de Lisboa: estudo de um grupo de indivíduos com elevado risco de infecção por vírus de transmissão parentérica.....	103
Caracterização genética das estirpes virais encontradas de GBV-C.....	106
Ensaio de PCR em Tempo Real - quantificação relativa do genoma de GBV-C.....	122
5. CONCLUSÕES E PERSPECTIVAS FUTURAS.....	127
6. REFERÊNCIAS BIBLIOGRÁFICAS.....	130
7. ANEXOS	
Anexo I.....	156
Anexo II.....	160
Anexo III.....	162
Anexo IV.....	164
Anexo V.....	167
Anexo VI.....	169
Anexo VII.....	172

ÍNDICE DE FIGURAS

1. INTRODUÇÃO

Figura 1.1.	Representação esquemática da partícula viral de GBV-C	6
Figura 1.2.	A) Organização genética e processamento da poliproteína do vírus da hepatite C; B) Representação esquemática da estrutura genómica do vírus GBV- C.....	7
Figura 1.3.	Representação simplificada do ciclo replicativo do HCV.....	10
Figura 1.4.	Esquema simplificado representativo da fusão do HIV à célula.....	16
Figura 1.5.	Múltiplos mecanismos propostos para a interferência de GBV-C no organismo hospedeiro numa situação de co-infecção com HIV.....	19
Figura 1.6.	Representação esquemática de putativas interações entre os vírus GBV-C, HCV, HIV e a doença hepática.....	20
Figura 1.7.	Representação esquemática da distribuição geográfica do vírus GBV-C.....	22
Figura 1.8.	Principais vias de transmissão do vírus GBV-C.....	25
Figura 1.9.	Representação das taxas de prevalência de GBV-C entre os dadores de sangue distribuídas globalmente.....	26

2. MATERIAL E MÉTODOS

Figura 2.1.	Caracterização da população estudada.....	32
Figura 2.2.	Caracterização da população estudada no que respeita à nacionalidade dos indivíduos envolvidos.....	33
Figura 2.3.	Esquema representativo da localização aproximada dos <i>primers</i> utilizados no decurso deste trabalho.....	38
Figura 2.4.	Esquema representativo do procedimento efectuado para purificação dos produtos de PCR com dois sistemas comerciais	43
Figura 2.5.	Esquema representativo do vector pGEM®-T.....	45
Figura 2.6.	Representação esquemática de um método de PCR em Tempo Real baseado na utilização de <i>SYBR Green I</i>	49
Figura 2.7.	Representação esquemática da sonda <i>TaqMan</i>	50

3. RESULTADOS

Figura 3.1.	Observação dos produtos de PCR resultantes da amplificação da região 5'UTR em gel de agarose.....	59
Figura 3.2.	Representação gráfica do número total de indivíduos, onde foi detectada a presença de RNA de GBV-C.....	60
Figura 3.3.	Representação gráfica do número de indivíduos, onde foi detectada a presença de co-infecções com HIV e HCV, e monoinfecção com GBV-C .	61
Figura 3.4.	Observação em gel de agarose dos produtos relativos à amplificação por PCR da região 5'UTR/E1 do genoma de GBV-C.....	63
Figura 3.5.	Exemplo de um dos cromatogramas que resultou da sequenciação das amostras com o <i>primer</i> E1rev4.....	64
Figura 3.6.	Alinhamento múltiplo das sequências referentes às amostras 9709, 1642, 821, 3273, 594 e 088 com a sequência de referência NC_001710.....	65
Figura 3.7.	Representação esquemática das estruturas secundárias obtidas para a região 5'UTR de GBV-C em estudos anteriores.....	66
Figura 3.8.	Representação gráfica das estruturas secundárias obtidas para as regiões HG6 (A) e HG7 (B) por analogia com o estudo de Thurner e colaboradores.....	67
Figura 3.9.	Representação esquemática do genoma de GBV-C.....	68
Figura 3.10.	Observação, em gel de agarose, dos produtos referentes à amplificação, por <i>nested</i> -PCR, da região E1/E2 e NS5A/NS5B.....	69
Figura 3.11.	A) Análise do padrão de migração, em gel de agarose, do DNA plasmídico de clones bacterianos; B) Perfil de digestão resultante da hidrólise, pela enzima EcoRI.....	71
Figura 3.12.	Exemplo de cromatogramas relativos às sequências nucleotídicas obtidas por sequenciação directa do produto de amplificação NS5A/NS5B.....	72
Figura 3.13.	Distribuição (percentagem) das estirpes de GBV-C em estudo, agrupadas em <i>clusters</i> monofiléticos pela análise da região E1/E2.....	75
Figura 3.14.	Relações filogenéticas entre as estirpes de GBV-C em estudo, baseadas na análise das sequências nucleotídicas da região E1/E2 (Árvore NJ).....	76
Figura 3.15.	Relações filogenéticas, baseadas na análise das sequências nucleotídicas obtidas para a região E1/E2 (Análise Bayesiana e critério ML).....	78

Figura 3.16.	Representação gráfica da análise obtida pela ferramenta PCOORD, para os grupos correspondentes ao genótipo 1, 2 e G2*.....	79
Figura 3.17.	Análise de <i>bootscanning</i> por <i>SimPlot</i> da sequência 6388E2.....	81
Figura 3.18.	Representação gráfica das distâncias genéticas inter e intragenótipos, considerando os genótipos 1,2,3 e 5 de GBV-C e o grupo G2* vs G2.....	82
Figura 3.19.	Representação gráfica das distâncias genéticas inter e intragenótipos, considerando os genótipos 1,2,3 e 5 de GBV-C e o grupo G2*.....	83
Figura 3.20.	Representação gráfica do efeito acumulativo da ocorrência de substituições sinónimas e não sinónimas por codão (E1/E2).....	84
Figura 3.21.	Representação gráfica das posições conservadas dos resíduos de aminoácidos (E1/E2).....	85
Figura 3.22.	Relações filogenéticas baseadas na análise das sequências nucleotídicas da região NS5A/NS5B (árvores de NJ).....	88
Figura 3.23.	Relações filogenéticas baseadas na análise das sequências nucleotídicas da região NS5A/NS5B (por análise Bayesiana e pelo critério de ML).....	90
Figura 3.24.	Análise de <i>bootscanning</i> por <i>SimPlot</i> (amostras 8227 e 1641).....	92
Figura 3.25.	Análise de <i>bootscanning</i> por <i>SimPlot</i> (amostras 1631 e 6388).....	93
Figura 3.26.	A) Representação gráfica da percentagem de estirpes de GBV-C em estudo, amplificadas para a região NS5A/NS5B, agrupadas com os respectivos genótipos; B) Análise obtida pela ferramenta PCOORD.....	94
Figura 3.27.	Comparação das distâncias filogenéticas, entre os grupos G2 ^a , G2* e G2.....	95
Figura 3.28.	Representação gráfica do efeito acumulativo da ocorrência de substituições sinónimas e não sinónimas por codão (NS5A/NS5B).....	96
Figura 3.29.	Representação gráfica das posições conservadas dos resíduos de aminoácidos (NS5A/NS5B).....	97
Figura 3.30.	A) Amplificação por PCR em Tempo Real de um fragmento de DNA (da 5'UTR); B) Representação gráfica das rectas de amplificação obtidas a partir dos valores de Ct.....	99
Figura 3.31.	Cromatograma respectivo à sequenciação da amostra com indicação das bases que revelaram posições nucleotídicas não homólogas com a sequência nucleotídica da sonda <i>TaqMan</i>	101

ÍNDICE DE TABELAS

1. INTRODUÇÃO

Tabela 1.1.	Principais grupos de risco associados a infecção por GBV-C.....	27
--------------------	-----------------------------------------------------------------	----

2. MATERIAL E MÉTODOS

Tabela 2.1.	Condições de reacção utilizadas nas experiências de síntese de cDNA.....	36
Tabela 2.2.	Informação relativa ao conjunto de <i>primers</i> delineados para amplificação de fragmentos de DNA de GBV-C por <i>nested</i> -PCR.....	37
Tabela 2.3.	Condições de reacção utilizadas na amplificação da região 5'UTR de GBV-C por <i>nested</i> -PCR.....	39
Tabela 2.4.	Condições de reacção utilizadas na primeira e segunda reacções de PCR destinada à amplificação da região 5'UTR/E1.....	40
Tabela 2.5.	Condições de reacção utilizadas nas reacções de PCR destinadas à amplificação das regiões E1/E2 e NS5A/NS5B.....	41
Tabela 2.6.	Condições de reacção utilizadas nas reacções de PCR destinadas à amplificação dos fragmentos de DNA clonados.....	48
Tabela 2.7.	Condições de reacção utilizadas na reacção de PCR em Tempo Real com o corante <i>SYBR Green I</i> , destinada à amplificação do DNA plasmídico correspondente à região 5'UTR.....	51
Tabela 2.8.	Informação relativa ao conjunto de <i>primers</i> e sonda usados para amplificação de fragmentos de DNA de GBV-C por PCR em Tempo Real.....	52

3. RESULTADOS

Tabela 3.1.	A) Posição nucleotídica do genoma de GBV-C (sequência NC_001710) onde terminaram precocemente cinco amostras em análise. B) Reconhecimento de potenciais estruturas secundárias na região E1 e 5'UTR de GBV-C.....	66
Tabela 3.2.	Identificação dos fragmentos clonados.....	70

Tabela 3.3.	Cálculo das distâncias genéticas entre as amostras indicadas e o respectivo grupo/genótipo em análise.....	96
Tabela 3.4.	Análise semi-quantitativa do número de cópias de DNA por reacção estimados a partir dos declives das respectivas rectas padrão.....	100

ABREVIATURAS

A,C,G,T	Bases orgânicas constituintes dos nucleótidos (adenina, citosina, guanina e timina, respectivamente)
BSA	Albumina sérica bovina
BSL	<i>Bio<u>S</u>afety <u>L</u>evel</i>
°C	Graus Celsius
CAP	<i><u>C</u>ontig <u>A</u>ssembly <u>P</u>rogram</i>
CDC	<i>Centers for Disease Control and Prevention</i>
cDNA	DNA complementar
DEPC	Dietil pirocarbonato
<i>d_N</i>	Substituições não sinónimas
DNA	Ácido desoxirribonucleico, do inglês <i><u>D</u>eoxyribo<u>n</u>ucleic <u>A</u>cid</i>
dNTP	Desoxirribonucleótidos 5'-trifosfatados
<i>d_S</i>	Substituições sinónimas
<i>E. coli</i>	<i>Eschericchia coli</i>
E1	Glicoproteína do invólucro 1, do inglês <i><u>E</u>nvelope glicoprotein <u>1</u></i>
E2	Glicoproteína do invólucro 2, do inglês <i><u>E</u>nvelope glicoprotein <u>2</u></i>
EDTA	Etilenodiaminotetracetato
EUA	Estados Unidos da América
FDA	<i>World <u>F</u>ood and <u>D</u>rug <u>A</u>dministration</i>
<i>g</i>	Aceleração da gravidade
GB	Iniciais do cirurgião G. Barker
GBV-A	Vírus GB-A, do inglês <i><u>G</u>B <u>V</u>irus <u>A</u></i>
GBV-B	Vírus GB-B, do inglês <i><u>G</u>B <u>V</u>irus <u>B</u></i>
GBV-C	Vírus GB-C, do inglês <i><u>G</u>B <u>V</u>irus <u>C</u></i>
G1-G6	Genótipos de 1 a 6 respectivamente
h	Hora
HAART	<i><u>H</u>ighly <u>A</u>ctive <u>A</u>nti<u>R</u>etroviral <u>T</u>reatment</i>
HAV	Vírus da hepatite A, do inglês <i><u>H</u>epatitis <u>A</u> <u>V</u>irus</i>
HBV	Vírus da hepatite B, do inglês <i><u>H</u>epatitis <u>B</u> <u>V</u>irus</i>
HBsAg	Antigénio B (HBV)
HCl	Ácido clorídrico
HCV	Vírus da hepatite C, do inglês <i><u>H</u>epatitis <u>C</u> <u>V</u>irus</i>
HDV	Vírus da hepatite D, do inglês <i><u>H</u>epatitis <u>D</u> <u>V</u>irus</i>
HEV	Vírus da hepatite E, do inglês <i><u>H</u>epatitis <u>E</u> <u>V</u>irus</i>

HIV-1	Vírus da imunodeficiência humana tipo 1, do inglês <i><u>H</u>uman <u>I</u>mmunodeficiency <u>V</u>irus type 1</i>
HIV-2	Vírus da imunodeficiência humana tipo 2, do inglês <i><u>H</u>uman <u>I</u>mmunodeficiency <u>V</u>irus type 2</i>
HFV	Vírus da hepatite F, do inglês <i><u>H</u>epatitis <u>F</u> <u>V</u>irus</i>
HGV	Vírus da hepatite G, do inglês <i><u>H</u>epatitis <u>G</u> <u>V</u>irus</i>
IDUs	<i><u>I</u>njecting <u>D</u>rug <u>U</u>sers</i>
IL	Interleucina
IPTG	Isopropil-β-D-tiogalactopiranosídeo
IRES	<i><u>I</u>nternal <u>R</u>ibosome <u>E</u>ntry <u>S</u>ite</i>
kDa	Kilo(Daltons) = 10 ³ Daltons
k(p)b	Kilo(pares) de bases = 10 ³ (pares) de bases
K2P	Kimura a dois parâmetros
LB	Meio de Luria-Bertani
M	Molaridade
MgCl₂	Cloreto de magnésio
MgSO₄	Sulfato de magnésio
m	Mili, prefixo que designa 10 ⁻³
M	Molar
MACS	<i><u>M</u>ulticenter <u>A</u>IDS <u>C</u>ohort <u>S</u>tudy</i>
MCMC	<i><u>M</u>arkov <u>C</u>hain <u>M</u>onte <u>C</u>arlo</i>
min	Minuto
MIP-1α	<i><u>M</u>acrophage <u>I</u>nflammatory <u>P</u>rotein <u>1</u>α</i>
MIP-1β	<i><u>M</u>acrophage <u>I</u>nflammatory <u>P</u>rotein <u>1</u>β</i>
ml	Mililitro
ML	<i><u>M</u>aximum <u>L</u>ikelihood</i>
mM	Milimolar
mRNA	Ácido ribonucleico mensageiro, do inglês <i><u>M</u>essenger <u>R</u>ibonucleic <u>A</u>cid</i>
NaOH	Hidróxido de sódio
NANB	Hepatite “nãoA-nãoB”
NCBI	<i><u>N</u>ational <u>C</u>enter for <u>B</u>iotechnology <u>I</u>nformation</i>
ng	Nanograma
NJ	<i><u>N</u>eighbor-<u>J</u>oining</i>
nm	Nanómetro = 10 ⁻⁹ metros
nM	Nanomolar
nt	Nucleótido
ORF	<i><u>O</u>pen <u>R</u>eadng <u>F</u>rame</i>
P	Probabilidade

pb	Pares de bases
PBMC	Células mononucleadas do sangue periférico, do inglês <i><u>P</u>eripheral <u>B</u>lood <u>M</u>ononuclear <u>C</u>ells</i>
PCR	Reacção de polimerização em cadeia, do inglês <i><u>P</u>olymerase <u>C</u>hain <u>R</u>eaction</i>
pH	Simétrico do logaritmo decimal da concentração hidrogeniónica de uma solução
poli-A	Cauda poliadenilada
pmol	Picomole = 10^{-12} moles
p/v	Massa de soluto por volume de solução
RANTES	<i><u>R</u>egulated on <u>A</u>ctivation <u>N</u>ormal <u>T</u> cell-<u>E</u>xpressed and <u>S</u>ecreted</i>
RNA	Ácido ribonucleico, do inglês <i><u>R</u>ibonucleic <u>A</u>cid</i>
RNase	Ribonuclease
RT	<i><u>R</u>everse <u>T</u>ranscription</i>
RT-PCR	Polimerização em cadeia da polimerase antecedida de transcrição reversa
SDF-1	<i><u>S</u>tromal-<u>D</u>erived <u>F</u>actor</i>
SDS	Dodecilsulfato de sódio, do inglês <i><u>S</u>odium <u>D</u>odecyl <u>S</u>ulphate</i>
seg	Segundos
SIDA	Síndrome da Imunodeficiência Adquirida
TAE	Tampão Tris-acetato EDTA
TE	Tampão Tris-EDTA
Th	Células <i>T-helper</i>
Tris	Tris(hidroximetil)aminometano
Tris-HCL	Tris Hidrocloreto
TSS	<i><u>T</u>ransformation and <u>S</u>torage <u>S</u>olution</i>
UDIs	<i><u>U</u>tilizadores de <u>D</u>rogas <u>I</u>njectáveis</i>
UTR	Região não traduzida, do inglês, <i><u>U</u>ntranslated <u>R</u>egion</i>
UV	Ultravioleta
vs	<i>versus</i>
μ	Micro, prefixo que designa 10^{-6}
μg	Microgramas
μl	Microlitros
μM	Micromolar

RESUMO



RESUMO

O vírus GBV-C é um dos potenciais membros da família *Flaviviridae* sendo, essencialmente, transmitido por via parentérica. Apesar de inicialmente descrito como agente de hepatites, dados recentes parecem demonstrar a não patogenicidade deste vírus. Curiosamente, o GBV-C parece exercer um efeito benéfico nos indivíduos co-infectados com HIV, facto que estimulou a realização de inúmeros estudos sobre a biologia deste vírus.

A escassez de informação de carácter genético/molecular relativa aos vírus GBV-C circulantes em Portugal, motivou a caracterização genética das estirpes virais circulantes na Grande Lisboa. Como corolário da abordagem utilizada, foi ainda possível avaliar a prevalência do GBV-C nesta área geográfica, tendo a análise efectuada recaído sobre um grupo de indivíduos, na sua maioria utilizadores de drogas injectáveis co-infectados com HIV e/ou HCV. A aplicação de um método para análise semi-quantitativa da carga viral de GBV-C constituiu, também, um dos propósitos deste trabalho.

Foram analisadas 214 amostras do soro humano nas quais o RNA de GBV-C foi detectado por uma reacção de transcrição reversa seguida de amplificação da 5'UTR por *nested*-PCR. A taxa de virémia na população estudada foi de 40,65%, tendo esta sido estatisticamente associada ao consumo de drogas e idade. As monoinfecções por GBV-C foram detectadas em 22,99% das amostras analisadas enquanto que a taxa de infecções duplas GBV-C/HIV ou GBV-C/HCV foram de 19,54% e 16,09%, respectivamente. Finalmente, as infecções triplas (GBV-C/HIV/HCV) foram registadas em 41,38% das amostras analisadas.

A caracterização de duas regiões distintas do genoma viral (E1/E2 e NS5A/NS5B) incluiu, entre outras, a análise filogenética das respectivas sequências nucleotídicas, tendo por base três modelos de reconstrução filogenética. A análise da região E1/E2 de 45 sequências virais, evidenciou a segregação de 10 destas com referências do genótipo 1 e 34 com referências do genótipo 2. Nestas últimas, um grupo de 16 sequências formou um putativo novo subtipo associado à área geográfica estudada, denominado G2*, e corroborado pela análise da região NS5A/NS5B.

A descoberta de novas variantes genómicas, a compreensão da sua distribuição, elucidação das suas características biológicas e interacção destas com outros vírus, dão ênfase à necessidade de continuidade dos estudos relacionados envolvendo o GBV-C.

ABSTRACT

The GB virus C (GBV-C) is a tentative member of the *Flaviviridae* family, readily transmitted by parenteral routes. Although initially regarded as a potential cause of hepatitis, more recent studies have disclosed no evidence for an association between GBV-C and human disease. With the demonstration of a positive effect of an active GBV-C infection on the outcome of HIV infection, the research on GBV-C has been recently reboosted.

Due to the scarcity of information regarding GBV-C infection in Portugal, we decided to carry out a genetic characterization of the viral strains circulating in the Greater Lisbon. An immediate consequence of this study was the evaluation of the prevalence of GBV-C infection in a population sample including a large number of IDUs, also characterized by high HIV and/or HCV seroprevalence. Furthermore, in order to quantify the GBV-C viral load, a semi-quantitative Real-Time-PCR analysis was also executed.

GBV-C viremia was assessed by nested-PCR amplification of a section of the conserved 5'-untranslated region, using total RNA extracted from 214 plasma samples. The overall prevalence of GBV-C infection was 40.6%. GBV-C viremia was found to be statistically associated with the age of the infected individuals or the use of intravenous drugs. Among those with GBV-C viremia, mono-infection was detected only in 22.99% of them, while co-infection with HIV, HCV, or HIV/HCV was revealed in 19.54%, 16.09%, and 41.38% of the population, respectively.

The genetic characterization of GBV-C of two distinct genomic regions (E1/E2 e NS5A/NS5B), was carried out by different approaches, including phylogenetic analysis of nucleotide sequences, based on three different methods. Of 45 viral strains comprising the E1/E2 genomic region, 10 showed segregation with genotype 1 and 34 with genotype 2 GBV-C references. However, 16 of the strains assigned to genotype 2 were shown to form a separate cluster (designated G2*), also confirmed by NS5A/NS5B genomic region analysis.

The discovery of newly genomic variants, the understanding of their geographical distribution, elucidation of biological characteristics and their interaction with other viruses, emphasize the need to keep on scientific research concerning GBV-C.

1. INTRODUÇÃO



1. INTRODUÇÃO

1.1. Nota introdutória

Existe um elevado número de vírus que, no decurso de uma infecção sistémica, concomitantemente com o envolvimento de outros órgãos, podem afectar o fígado, originando uma hepatite. Em muitos casos a hepatite não chega a ter expressão clínica, sendo detectada, ocasionalmente, por alterações de carácter bioquímico. O termo hepatite vírica ou viral é, convencionalmente, atribuído a patologias causadas por um restrito subgrupo de vírus designados hepatotrópicos, cuja descoberta foi responsável pelo extraordinário progresso verificado, nas últimas décadas, no domínio da hepatologia. As hepatites virais são consideradas um problema de saúde humana à escala global, com elevado impacto na morbilidade e mortalidade das populações infectadas, chegando a ser consideradas a maior pandemia mundial da actualidade (CDC, *Center for Disease Control and Prevention*, disponível em <http://www.cdc.gov/hepatitis/>).

Após a descoberta do antígeno de superfície do vírus da hepatite B (HBV, *Hepatitis B Virus*) em 1965 (Blumberg *et al.*, 1965), e do vírus da hepatite A (HAV, *Hepatitis A Virus*) em 1973, (Feinstone *et al.*, 1973) ficaram bem definidos dois dos principais agentes responsáveis pelas hepatites virais. Mais tarde, em 1977, e no decurso de estudos, por imunofluorescência, de biopsias hepáticas provenientes de indivíduos infectados com o vírus da hepatite B (HBV), um grupo de investigadores identificou um novo antígeno, imunologicamente distinto de qualquer um dos antígenos do HBV (Rizzetto *et al.*, 1977), o qual foi provisoriamente apelidado de Delta, correspondendo a designação anti-Delta à do(s) anticorpo(s) correspondente(s) circulante(s). Posteriormente, este agente viria a ser reconhecido como vírus da hepatite D (HDV, *Hepatitis D Virus*), estando a sua replicação e expressão directamente associadas ao HBV (Polish, *et al.*, 1993), do qual é um vírus satélite.

A introdução de técnicas cada vez mais sensíveis de detecção do antígeno B (HBsAg) foi progressivamente conduzindo à redução do risco de transmissão pós-transfusional do HBV. Apesar disso, na década de 1970, os casos de hepatite continuaram a ocorrer entre 5 a 10% dos doentes que recebiam sangue ou seus derivados (Feinstone *et al.*, 1975). Ainda que o HAV tenha sido inicialmente apontado como o agente etiológico responsável por uma fracção considerável dos casos de hepatite pós-transfusional, os resultados de estudos epidemiológicos, combinados com a introdução de novas técnicas laboratoriais, tornaram evidente que o HAV não estava

presente em nenhum dos casos. Como tal, as hepatites em questão adquiriram a designação de hepatite “nãoA-nãoB” (ou NANB). Esta denominação, introduzida em 1975, viria a manter-se durante 14 anos, até a identificação do vírus da hepatite C (HCV, *Hepatitis C Virus*), apontado como o agente viral associado ao maior número de casos de hepatite pós-transfusional NANB (Choo *et al.*, 1989; Alter, 1999). Para além deste vírus de transmissão parentérica, uma hepatite de tipo NANB de natureza epidémica, associada a um agente de transmissão fecal-oral, geralmente benigna e, normalmente, sem evolução para a cronicidade, foi designada de hepatite E. O vírus por ela responsável (o vírus da hepatite E ou HEV, *Hepatitis E Virus*) foi identificado um ano mais tarde (Aggarwal e Krawczynski, 2000). Todos os quatro vírus até então descritos, causam hepatite aguda, mas apenas as hepatites B, C e D podem evoluir para um quadro de hepatite crónica (60 a 80% dos casos num espaço de 15-20 anos), podendo, entre 1 a 2% das infecções, progredir para cirrose hepática ou hepatocarcinoma (Cheung *et al.*, 1997; CDC, disponível em <http://www.cdc.gov/hepatitis/>).

Com a identificação do HCV em 1989, e do HEV em 1990, a expressão hepatite NANB parecia já não fazer sentido. No entanto, cedo se verificou que um apreciável número de casos de hepatite de etiologia viral e de tipo NANB, não era estava relacionado com os vírus HCV ou HEV, o que obrigou a uma modificação da designação clínica para hepatite “não A-E” (Karayianni e Thomas, 1997). A evidência da existência de um ou mais vírus relacionados com o desenvolvimento destas hepatites adveio do facto de que aproximadamente 10% das hepatites pós-transfusionais, 20% das hepatites adquiridas na comunidade, 30% dos casos de hepatite crónica e cirrose criptogénica, e uma larga percentagem de casos de hepatites fulminantes, não eram causados por qualquer dos vírus hepatotrópicos conhecidos (Ferraz, *et al.*, 1996; Freitas, 2003).

Em Dezembro de 1994 um grupo de investigadores reportou o isolamento de partículas virais com 27-37nm de diâmetro, nas fezes de doentes com hepatites pós-transfusional “não A-E” (Deka *et al.*, 1994). A inoculação destas partículas em macacos *Reshus* resultou no desenvolvimento de hepatite, e o vírus identificado passou a ser denominado por vírus da hepatite F (HFV, *Hepatitis F Virus*) ou Toga vírus (Deka *et al.*, 1994). Contudo, estudos relativamente recentes falharam na confirmação da existência de uma associação entre a presença deste vírus e a hepatite humana, razão pela qual foi retirado da lista de hepatites infecciosas (Bowden, 2001; Kelly e Skidmore, 2002).

O impacto dos vírus causadores de hepatite na saúde humana tem motivado, ao longo do tempo, inúmeros estudos que visam a descoberta/caracterização de novos agentes virais hepatotrópicos.

Curiosamente, e ainda que nos últimos 10 anos tenham sido vários os vírus inicialmente descritos como potenciais agentes etiológicos de hepatite pós-transfusional (de entre os quais se destacam o vírus TT, o vírus SEN ou o GBV-C), as suas consequências na saúde humana permanecem ainda por esclarecer (Williams e Riordan, 2000; Bowden, 2001; Wiwanitkit, 2006).

1.2. Os vírus GB: uma perspectiva histórica

O desencadeamento da sucessão de eventos que levou à descoberta do vírus GB-C (GBV-C, *GB Virus C*) teve início em 1967, data em que foi publicada uma tentativa bem sucedida de transmissão de hepatite viral humana a primatas do Novo Mundo (dos géneros *Saguinus* e *Callithrix*), mediante a utilização de inóculos de soro de doentes com presumível hepatite pós-transfusional “não A-E” (Deinhardt *et al.*, 1967). Em particular, alíquotas de soro retiradas de um cirurgião, de seu nome G. Barker (GB), de 34 anos, originário de Chicago, e que apresentava uma hepatite aguda de origem desconhecida, foram injectadas em 4 macacos tamarinos (género *Saguinus*) nos quais se verificou o desenvolvimento de hepatite entre 16-40 dias após injeção. O potencial infeccioso do(s) agente(s) viral(ais) em causa, foi mantido em passagens sucessivas de amostras de soro nos tamarinos até, pelo menos, à 11ª passagem (amostra designada H205 GB pass11) (Simons *et al.*, 1995). Os diferentes testes laboratoriais efectuados permitiram demonstrar que a hepatite desenvolvida pelo cirurgião GB não fora causada pelos vírus HBV, HCV ou HEV (Deinhardt *et al.*, 1967; Bowden, 2001), tornando-se, deste modo, claro, que a amostra H205 GB pass11 era uma fonte potencial de um *novo* vírus de hepatite, até então não caracterizado. Este veio a ser designado de agente GB, de acordo com as iniciais da sua suposta fonte (Deinhardt *et al.*, 1975). Embora tenham sido realizados ensaios sucessivos para isolar e identificar o vírus GB através da aplicação de métodos clássicos, a sua completa caracterização só se tornou possível com o advento da biologia molecular, nomeadamente no que se refere ao desenvolvimento da técnica de RT-PCR (polimerização em cadeia da polimerase antecedida de transcrição reversa) (Dawson *et al.*, 1996). As tentativas de caracterização começaram, por conseguinte, 25 anos mais tarde quando em 1995, usando a técnica de Análise de Diferenças Representativas (*Representation Differencel Analysis*), investigadores dos laboratórios Abbott identificaram no soro H205 GB pass11 dois genomas virais. Estes demonstraram estar relacionados com os genomas dos vírus da família *Flaviviridae*, tendo vindo a ser denominados por GBV-A (*GB Virus A*) e GBV-B (*GB Virus B*) (Simons *et al.*, 1995). Para além disso, foi ainda evidenciado o facto de apenas o vírus GBV-B ter capacidade para causar hepatite em

macacos tamarinos, podendo, como tal, representar o agente GB original (Schlauder *et al.*, 1995). Por outro lado, e embora tenha sido demonstrado que a infecção somente com GBV-A não resultava no desenvolvimento de hepatite, pensa-se que a co-infecção com ambos os vírus GBV-A e GBV-B, conduza a uma hepatite com grau de severidade superior relativamente a situações de monoinfecção com GBV-B (Bowden, 2001). No entanto, ainda que na população em risco para hepatite “não A-E” se verifique a presença de anticorpos reactivos contra antígenos de GBV-A e GBV-B (Simons *et al.*, 1995; Pilot-Matias *et al.*, 1996), ambos são considerados vírus de primatas não-humanos (Muerhoff *et al.*, 1995; Erker^a *et al.*, 1998; Chavez *et al.*, 2009).

O mesmo grupo rapidamente desenvolveu testes serológicos baseados na obtenção de proteínas recombinantes dos vírus GBV-A e GBV-B para estudos de seroprevalência. Soro humano de indivíduos considerados em risco de infecção com vírus causadores de hepatite “não A-E”, foram então rastreados e, dos soros imuno-reactivos, foi extraído RNA total, no qual se viria a pesquisar a presença de sequências virais. Para superar a possível variação genética, foram usados *primers* degenerados, complementares a regiões conservadas do putativo gene que codifica a helicase (NS3) dos vírus GBV-A, GBV-B e HCV. Na sequência dos estudos efectuados, viria a ser identificado um terceiro vírus, o qual foi associado ao grupo GBV dado o grau de semelhança antigénica com o vírus GBV-B (revelado em ensaios imunológicos), e proximidade filogenética com o vírus GBV-A (Leary *et al.*, 1996), pelo que lhe foi atribuída a designação de GBV-C (Simons *et al.*, 1995; Deka *et al.*, 1994). Simultaneamente, um dos grupos de trabalho da *Genelab Technologies*, em colaboração com o *Center for Disease Control and Prevention* (CDC, EUA), e mediante a utilização de uma estratégia baseada na detecção de antígenos virais expressos a partir de um banco genómico (abordagem experimental semelhante à que, anos antes, tinha permitido a descoberta do HCV), conduziu à detecção de uma sequência homóloga à de alguns flavivírus em amostras de doentes com hepatite NANB. Esta foi, provisoriamente, associada à presença de um vírus nessas amostras, o qual veio a ser designado de vírus da hepatite G ou HGV (*Hepatitis G Virus*) (Linnen *et al.*, 1996).

Quando a sequência genómica de ambos os vírus (GBV-C e HGV), se tornou disponível verificou-se que partilhavam um elevado grau de semelhança ao nível nucleotídico e proteico (86% e 95%, respectivamente), indicando corresponderem a estirpes do mesmo vírus, e razão pela qual este passou a ser conhecido como GBV-C/HGV (Polgreen *et al.*, 2003). O vírus é então referido actualmente na literatura com a designação taxonómica GBV-C/HGV (Polgreen *et al.*, 2003). No entanto, e curiosamente, este parece não estar associado ao desenvolvimento de

hepatite (ver secções seguintes), razão pela qual, a maioria dos grupos de investigação atribui a denominação apenas de GBV-C, também adoptada nesta dissertação.

1.3. O vírus GBV-C

1.3.1. Características gerais da partícula viral e estrutura genómica

A caracterização molecular do vírus GBV-C permitiu a sua possível inclusão, na família *Flaviviridae*. Enquanto membro desta família de vírus, o GBV-C partilha algumas características com os restantes flavivírus, nomeadamente no que se refere à organização da partícula viral e do seu genoma (Simmonds^b, 2001; Stapleton, 2003). Com efeito, o virião de GBV-C é limitado por um invólucro constituído por uma bicamada lipídica onde estão inseridas as glicoproteínas de superfície E1 e E2 (Reshetnyak *et al.*, 2008). O invólucro limita internamente uma putativa nucleocápside (Xiang *et al.*, 1998), que contém o genoma viral (Figura 1.1).

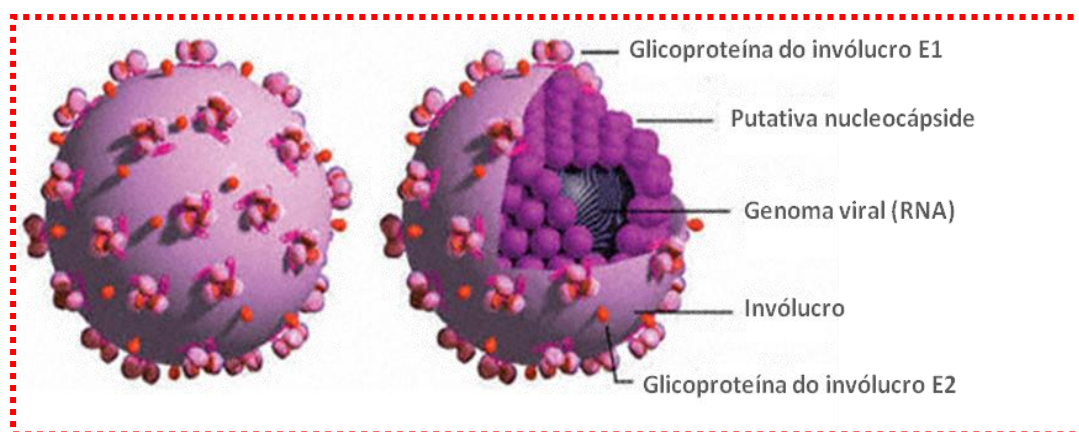


Figura 1.1. Representação esquemática da partícula viral de GBV-C (adaptado de http://www.hivandhepatitis.com/hiv_hcv_co_inf/2009/010909_a.html).

O genoma deste vírus é constituído por uma única cadeia linear de RNA de polaridade positiva com aproximadamente 9400 nucleótidos (Leary *et al.*, 1996; Okamoto *et al.*, 1997; Polgreen *et al.*, 2003). Este genoma contém uma única grelha de leitura aberta (ORF, *Open Reading Frame*) que codifica uma única poliproteína com 2873-2910 resíduos de aminoácidos, a qual, à semelhança do que acontece com inúmeros outros flavivírus, deverá ser subsequentemente clivada por proteases celulares e virais, de forma a originar as diferentes proteínas virais maduras (Reshetnyak *et al.*, 2008). A região codificante do genoma viral encontra-se precedida por uma

região não traduzida (5'UTR, *Untranslated Region*) com cerca de 460 nucleótidos (Viazov *et al.*, 1997), sendo esta, a região mais conservada do genoma viral. Esta região do genoma de GBV-C contém um sítio de entrada interno para ribossomas (IRES, *Internal Ribosome Entry Site*), o qual assegura a tradução da poliproteína viral. No extremo oposto do genoma, a região codificante é limitada por uma segunda região não traduzida (3'UTR) a qual deverá estar implicada na regulação da transcrição viral (Simons *et al.*, 1996; Bassit *et al.*, 1998).

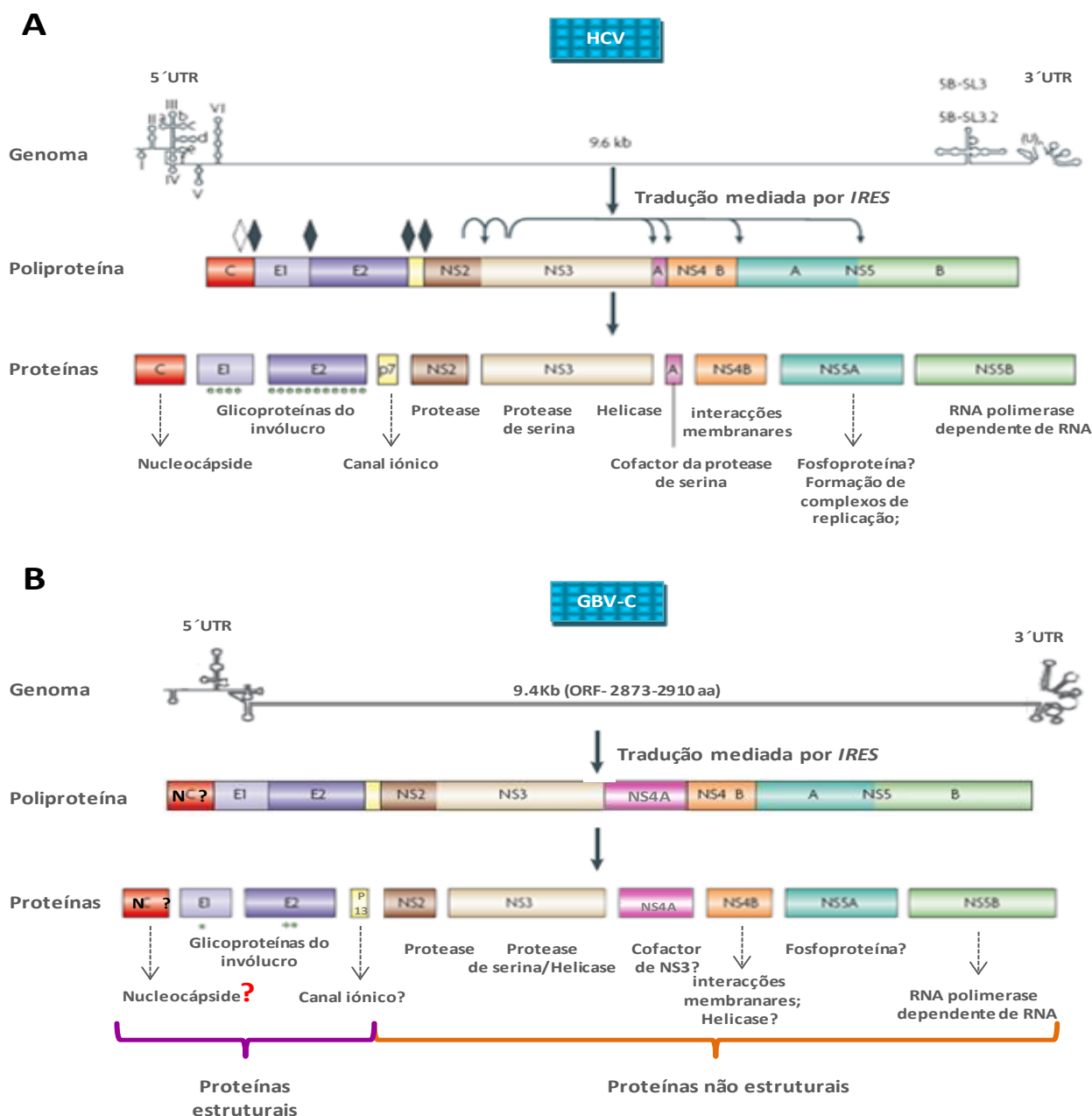


Figura 1.2. A) Organização genética e processamento da poliproteína do vírus da hepatite C (adaptado de Chevaliez e Pawlotsky, 2006; Moradpour *et al.*, 2007; Suzuki *et al.* 2007). B) Representação esquemática da estrutura genómica do vírus GBV-C, putativas proteínas codificadas e função associada. A existência de uma estrutura codificante para uma hipotética nucleocápside encontra-se definida a vermelho e pelo caractere “?”.

Pela análise da estrutura genómica de ambos os vírus GBV-C e HCV (Figura 1.2 A e B), torna-se evidente o grau de semelhança relativo à organização da mesma, onde os genes estruturais estão posicionados na metade 5' do genoma, enquanto que os genes que codificam as proteínas não estruturais localizam-se na extremidade oposta (Leary *et al.*, 1996; Simmonds^b, 2001; Stapleton, 2003).

As proteínas estruturais codificadas pelo RNA do GBV-C incluem as glicoproteínas do invólucro, designadas E1 e E2 (*Envelope glicoproteins*). Pela análise da sequência das regiões que codificam estas glicoproteínas verificou-se a inexistência de hipervariabilidade nas mesmas (em particular na E2), ao contrário do que acontece com as regiões homólogas do HCV (Stapleton *et al.*, 2004). A análise comparativa das sequências dos genomas do GBV-C e HCV sugere ainda a existência de uma região que codifica uma pequena proteína com 13kDa localizada entre E2 e NS2 (Ghibaud *et al.*, 2004) a qual deverá participar na formação de canais iónicos e, possivelmente, estará ainda implicada na morfogénese dos viriões e/ou na entrada dos vírus na célula hospedeira.

As proteínas NS2 e NS3 deverão compor a protease viral e, para além disso, NS3 deverá ter também actividade de helicase de RNA. NS4A poderá ser um cofactor da proteína NS3, ao passo que NS4B é uma proteína membranar de função ainda não totalmente conhecida (Lundin *et al.*, 2003). NS5A deverá participar na regulação da resposta imune do hospedeiro na formação dos complexos de replicação (à semelhança do que tem sido descrito para o HCV) e, por último, a proteína NS5B corresponde a uma polimerase de RNA dependente de RNA. As proteínas referidas são consideradas proteínas não estruturais (Stapleton *et al.*, 2004; Berzsenyin^a *et al.*, 2005; George e Varmaz, 2005; Chevaliez e Pawlotsky, 2006; Moradpour *et al.*, 2007; Suzuki *et al.*, 2007; Reshetnyak *et al.*, 2008; Stapleton *et al.*, 2009).

Ainda que existam semelhanças entre as hipotéticas proteínas codificadas pelo genoma do GBV-C e as codificadas pelo HCV, estas estão, normalmente, limitadas a motivos enzimáticos específicos, como os identificados nas proteínas NS3 e NS5B. A nível das regiões situadas nas extremidades do genoma, a região 5'UTR do GBV-C (≈460pb) é maior do que a mesma região no HCV (apenas com ≈340pb). Por outro lado, a extremidade 3'UTR (≈300pb) não contém a cauda poliadenilada (poli-A), ou uma região com poli-U, ao contrário do que sucede no HCV (Stapleton *et al.*, 2009), a qual se localiza entre o final da grelha de leitura e os últimos nucleótidos do extremo 3' do genoma. A semelhança entre os produtos de ambos os vírus é ainda menos evidente quando analisadas as glicoproteínas de superfície, para as quais se verificou existir uma reduzida identidade entre ambos os vírus (Simmonds^a, 2001). Para além da atenuada

semelhança ao nível da sua sequência primária, enquanto que as proteínas do invólucro do HCV são altamente glicosiladas (15 locais de glicosilação: 4 para a região E1 e 11 na proteína E2) as proteínas E1 e E2 do GBV-C têm glicosilação limitada, tendo sido identificados apenas 1 sítio de glicosilação na proteína E1 e 2 potenciais sítios de glicosilação na E2 (Muerhoff *et al.*, 1996; Baggio-Zappia e Granato 2009).

Apesar dos inúmeros estudos até hoje efectuados, várias são as questões relativas à caracterização da partícula viral do GBV-C que permanecem sem resposta. Exemplo dessa situação, é o da incapacidade de caracterização inequívoca da estrutura da nucleocápside viral, ou mesmo da identificação, no genoma viral, de uma região que a possa codificar (Okamoto *et al.*, 1997; Polgreen *et al.*, 2003; Marmor, 2006). Embora a presença de uma putativa nucleocápside viral, permaneça, até hoje, desconhecida, a sua observação por microscopia electrónica em partículas de GBV-C foi já sugerida (George e Varmaz, 2005).

1.3.2. Tropismo e replicação do vírus GBV-C

O GBV-C pode replicar em células mantidas em cultura, embora o local ou locais onde ocorre replicação do GBV-C *in vivo* não foi(ram), até hoje, estabelecidos de modo conclusivo (George e Varmaz, 2005). Existem estudos documentados onde se observa a replicação do vírus, em baixos níveis, nas células hepáticas, tendo já sido cultivado *in vitro* tanto em linhagens de células derivadas de linfócitos (células Daudi, MT-2 e MOLT-4), como hepáticas (como Huh-7, HepG2, PH5CH) (Xiang *et al.*, 2000; Cao *et al.*, 2009). Apesar da detecção de RNA viral no fígado ou em células de hepatoma em cultura demonstrada por alguns autores (Saito *et al.*, 1997; Seipp *et al.*, 1999), estes resultados foram, no entanto, ambíguos quando comparados com outros trabalhos (Laskus *et al.*, 1998; Radkowski *et al.*, 2000; Pessoa *et al.*, 1998), tendo conduzido à hipótese de que o GBV-C não seja um vírus hepatotrópico como o HCV (Howard, 2002; Stapleton, 2003; Baggio-Zappia e Granato, 2009). Embora o tropismo viral não tenha sido, até à data, claramente determinado, o GBV-C parece ser capaz de infectar diferentes células hematopoiéticas ou de linhagem linfóide, tais como os linfócitos CD4⁺ (Reshetnyak *et al.*, 2008). Trabalhos realizados *in vivo* envolvendo linhagens de células mononucleadas do sangue periférico (PMBCs, *Peripheral Blood Mononuclear Cells*), sustentam a hipótese de que os PMBCs sejam, efectivamente, o primeiro local de replicação do vírus GBV-C (Xiang *et al.*, 2000; Xiang *et al.*, 2001). Esta conjectura, é ainda apoiada pela pronta detecção de RNA viral, não só em células mononucleadas do sangue periférico (principalmente nos linfócitos B e T

CD4⁺ e CD8⁺), como também nos nódulos linfáticos (Radkowski *et al.*, 1999), em células do endotélio vascular (Handa e Brown, 2000), no baço e medula óssea (Laskus *et al.*, 1998; Kao *et al.*, 1999; Handa e Brown, 2000; Radkowski *et al.*, 2000; Tucker *et al.*, 2000; Xiang *et al.*, 2000; Xiang *et al.*, 2001; George *et al.*, 2003). Assim, o GBV-C é considerado por muitos investigadores como um vírus panlinfotrófico (Tucker *et al.*, 2000; Stapleton, 2003; Reshetnyak *et al.*, 2008).

1.3.3. Ciclo replicativo

Não existem dados disponíveis sobre o ciclo replicativo do GBV-C. No entanto, presume-se que a sua estratégia replicativa seja semelhante à do HCV (esquemáticamente representado na Figura 1.3) (George e Varmaz 2005). Como tal, a entrada do GBV-C na célula hospedeira deverá decorrer da interacção entre as glicoproteínas de superfície e um ou mais receptores celulares ainda não identificados (Berzsenyi^a *et al.*, 2005; Mohr e Stapleton, 2009).

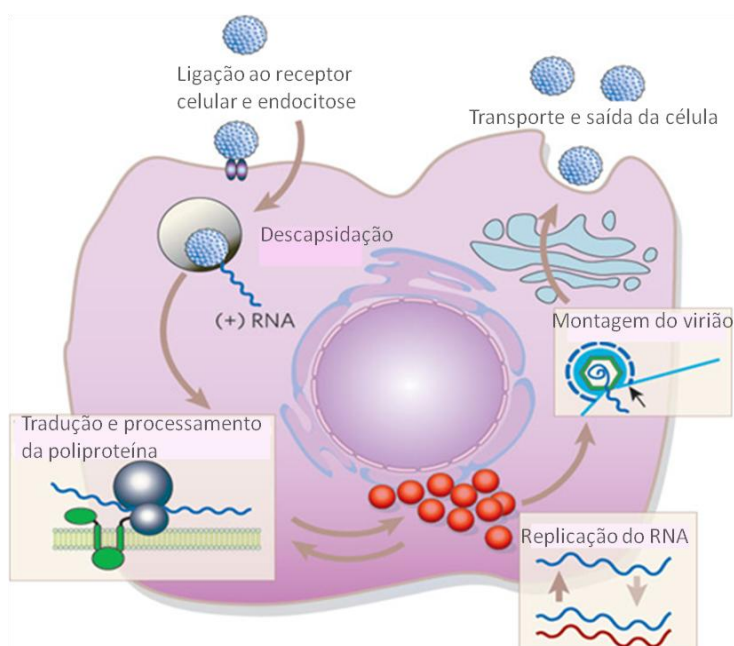


Figura 1.3. Representação simplificada do ciclo replicativo do HCV (Lindenbach *et al.*, 2007).

A penetração das partículas virais no citoplasma deverá ser sucedida da sua descapsidação, com consequente libertação do genoma viral. Dadas as suas características, este deverá ser traduzido de imediato dando origem a uma poliproteína viral, que posteriormente será processada por proteases celulares e virais. A replicação do genoma viral deverá ocorrer em associação a

membranas dando origem a genomas virais que deverão ser encapsidados (provavelmente a nível do retículo endoplasmático) contendo proteínas do invólucro sintetizadas *de novo* (George e Varmaz, 2005; Mohr e Stapleton, 2009). Ressalva-se o facto de nenhum destes processos ter sido directamente visualizado, sendo apenas pressupostos baseados na similitude com outros Flavivírus.

1.3.4. Infecção e possível associação a doença

A infecção com GBV-C pode resultar numa infecção persistente, a qual pode subsistir por mais de uma década, tendo infecções com mais de 16 anos de duração sido já documentadas (Kao *et al.*, 1997; Abraham, 2007; Reshetnyak *et al.*, 2008). No entanto, a grande maioria dos indivíduos infectados, consegue eliminar o vírus da circulação sanguínea (Thomas *et al.*, 1998; Williams *et al.*, 2004; George e Varmaz, 2005) podendo, ou não, desenvolver anticorpos contra a glicoproteína do invólucro E2, sendo a re-infecção esporádica (Devereux *et al.*, 1998) ou inexistente (Thomas *et al.*, 1998). Geralmente, a produção de anticorpos contra a proteína E2 do GBV-C (também denominados por anticorpos anti-GBV-C), é concomitante com o desaparecimento do RNA viral do soro (Thomas *et al.*, 1998; Gutierrez *et al.*, 1997; Feucht *et al.*, 1997), admitindo-se que estes anticorpos sejam aparentemente neutralizantes (Thomas *et al.*, 1998). Como tal, a presença de anticorpos anti-E2 indica a resolução de uma infecção activa, ou em alternativa, é caracterizada como um marcador de infecção passada (Reshetnyak *et al.*, 2008; Thomas *et al.*, 1998; Tillmann *et al.*, 1998). A exposição a GBV-C numa dada população pode, desta forma, ser atribuída à soma das taxas de detecção de RNA no plasma (virémia) e de anticorpos anti-E2, (Abraham, 2007; Stapleton *et al.*, 2004). A prevalência destes anticorpos no soro é 2 a 6 vezes superior à da detecção de virémia, sugerindo, tal como acima se indica, que a eliminação da infecção é comum (Thomas *et al.*, 1998; Berzsenyi^a, *et al.* 2005). Ainda que a resolução da infecção não implique, necessariamente, o desenvolvimento de anticorpos contra a proteína E2 (Williams *et al.*, 2004), o desenvolvimento de anti-E2 está também associado à protecção contra re-infecções (Tillmann *et al.*, 1998; Thomas *et al.*, 1998).

Na população saudável, a taxa de cronicidade subsequente a uma infecção aguda com GBV-C não é conhecida com precisão, embora tenham sido indicados valores situados entre os 25 e 50% (Alter *et al.*, 1997). Dada a capacidade que o GBV-C apresenta para estabelecer infecções crónicas persistentes que podem durar toda a vida dos indivíduos infectados, este vírus deverá

utilizar mecanismos de escape à resposta imune do hospedeiro infectado, os quais permanecem, no entanto, por esclarecer (Orii *et al.*, 2000).

Estima-se que o vírus GBV-C infecte, aproximadamente, 1/6 da população mundial (Muerhoff *et al.*, 2005; Reshetnyak *et al.*, 2008) com frequências de infecção activa, ou passada, a atingir valores entre os 5 e 15%, respectivamente (Simmonds^a, 2001). Apesar da elevada taxa de infecção, são escassos os episódios descritos na literatura que sugerem que o GBV-C esteja relacionado a doença hepática, apesar da persistência de virémia (Heringlake *et al.*, 1996; Alter *et al.*, 1997; Kubo *et al.*, 1997; Rey *et al.*, 2000). Na realidade, foram já vários os estudos que parecem indicar que a infecção pelo GBV-C, não está associada ao desenvolvimento de qualquer patologia específica. De facto, não foi reconhecida qualquer associação entre a infecção pelo GBV-C e o desenvolvimento de hepatite fulminante, hepatite aguda ou crónica (Alter *et al.*, 1997; Laskus *et al.*, 1997; Muñoz *et al.*, 1999), doença do fígado crónica pós-transplante hepático (Davison *et al.*, 1998), carcinoma hepatocelular (Kubo *et al.*, 1997; Tagger *et al.*, 1997), crioglobulinémia, ou desordens hematológicas de origem maligna e não maligna (Simmonds^a, 2001; Polgreen *et al.*, 2003).

Por outro lado, o GBV-C é frequentemente encontrado em situações de co-infecção com outros vírus, tais como os vírus da hepatite B e C (Mphahlele *et al.*, 1999; Barusruk *et al.*, 2006; Li^a *et al.*, 2006; Yang *et al.*, 2006; Zhang *et al.*, 2006; Tillmann *et al.*, 2008). No entanto a co-infecção com o GBV-C parece não agravar o curso das hepatites crónicas associadas ao HBV e/ou HCV (Alter *et al.*, 1997, Bjorkman, 2001). Em consequência, e apesar da elevada prevalência da infecção persistente com GBV-C neste grupo de indivíduos, a infecção pelo GBV-C não aparenta representar um risco substancial para a saúde humana, razão pela qual a organização FDA (*World Food and Drug Administration*) considera desnecessária a recomendação para triagens serológicas do GBV-C e/ou testes de virémia de rotina nos doadores de sangue (Reshetnyak *et al.*, 2008).

1.4. Impacto do GBV-C nas co-infecções com HIV e/ou com HCV

1.4.1. Relevância biológica

Tal como referido anteriormente, não foi, até à data, registada qualquer associação significativa entre a infecção pelo GBV-C e o desenvolvimento de qualquer patologia nos humanos. Por outro

lado, inesperadamente, e apesar de alguns relatos contraditórios na literatura (Bjorkman *et al.*, 2004; Stapleton *et al.*, 2004; Kaye *et al.*, 2005; Van der Bij *et al.*, 2005), têm sido vários os estudos que sugerem um efeito “protector” relativamente à progressão das infecções pelo vírus da Imunodeficiência Humana, HIV (*H*uman *I*mmunodeficiency *V*irus) nos indivíduos co-infectados com o GBV-C (Tillmann e Manns, 2001; Williams *et al.*, 2004; Zhang *et al.*, 2006; Hattori *et al.*, 2007; Handelsman *et al.*, 2007). O primeiro registo mencionando esta desaceleração, com consequente diminuição da taxa de mortalidade de co-infectados com HIV/GBV-C, data de 1998 (Toyoda *et al.*, 1998), tendo sido promotor de um interesse renovado no estudo do GBV-C (Baggio-Zappia e Granato, 2009).

Actualmente ainda se desconhece a forma como a diversidade genética de GBV-C afecta as interacções entre os dois vírus, ou se existe um impacto na progressão para SIDA nos doentes co-infectados, associada à infecção com o vírus GBV-C de algum genótipo específico (Muerhoff *et al.*, 2006; Berzsenyi *et al.*, 2009). Contrariando a anterior sugestão de que a variabilidade genotípica de GBV-C é reduzida, e que pouco impacto poderá revelar em termos fenotípicos (Simmonds e Smith 1999; Pavesi, 2001; Simmonds^a, 2001), estudos realizados com doentes seropositivos para HIV-1 (*H*uman *I*mmunodeficiency *V*irus Type 1) e co-infectados com GBV-C do genótipo 2 e 5, parecem evidenciar a existência de efeitos benéficos na progressão da doença dos doentes e, como tal, uma associação favorável resultante da co-infecção de ambos os vírus (George *et al.*, 2003; Muerhoff *et al.*, 2003; Sathar *et al.*, 2004; Muerhoff *et al.*, 2005). Assim, a sugestão de que a infecção pelo GBV-C poderá melhorar a sobrevivência nos doentes infectados pelo HIV, abre a porta a uma área de investigação no tratamento das infecções causadas por este vírus.

Na sequência da importância clínica que a replicação do GBV-C pode incutir na progressão das infecções pelo HIV, subsiste também um manifesto interesse na envolvente que resulta da co-infecção por GBV-C/HCV (Berzsenyi *et al.*, 2007; Barbosa *et al.*, 2009). Relativamente a co-infecções com outros vírus de hepatites, nomeadamente com o vírus da hepatite B, existem poucos trabalhos realizados neste âmbito, embora um estudo de 1998, tenha demonstrado não existir, aparentemente, um efeito significativo do GBV-C sobre os níveis de virémia de HBV em doentes co-infectados (Kao *et al.*, 1998). A co-infecção com GBV-C ocorre com uma prevalência de 10-25% em doentes com hepatite crónica C ou B, dependendo da população estudada (Petrik *et al.*, 1998; Barusruk *et al.*, 2006), embora infecções triplas com HIV/HCV/GBV-C tenham sido já registadas (Stapleton, 2007; Piroth *et al.*, 2008).

Os dados actuais indicam que a associação do GBV-C, com o HCV, ou em doentes co-infectados com HIV-HCV, não induz um aumento na progressão da doença hepática (Strauss *et al.*, 2002), sustentando o princípio de que o GBV-C é um vírus não-patogénico. Os escassos estudos efectuados que estão relacionados com a co-infecção GBV-C/HBV sugerem ainda que o GBV-C não interfere na replicação desse vírus (Baggio-Zappia e Granato, 2009).

1.4.2. Dinâmica da co-infecção GBV-C/HIV

A infecção com HIV está associada a uma panóplia de cenários clínicos, que vão desde uma infecção aguda inicial, muitas vezes inaparente, passando por uma fase de latência clínica, a qual é acompanhada da erosão progressiva do sistema imune, à desabilitação parcial, ou total, da resposta imunológica, na ausência de terapia anti-retroviral de alta eficácia (HAART), culminando no estadio de SIDA (Siliciano, 2005). Os primeiros trabalhos que relacionam a co-infecção com ambos os vírus sugeriam que a co-infecção com HIV e GBV-C proporcionava um quadro clínico favorável para os indivíduos infectados pelo HIV, na medida em que atrasava a progressão da infecção para o estadio de SIDA comparativamente com os doentes apenas monoinfectados por HIV (Lefrère^b *et al.*, 1999; Tillmann e Manns, 2001; Xiang *et al.*, 2001; Williams *et al.*, 2004). Nos indivíduos co-infectados observou-se uma redução na taxa de mortalidade, melhoria da qualidade de vida, redução da virémia relativa ao HIV e aumento do número de células T CD4⁺, independentemente de tratamento prévio com drogas de acção anti-retroviral, idade, raça/etnia, sexo, ou via de transmissão do vírus HIV, quando comparados com doentes mono-infectados (Yeo *et al.*, 2000; Xiang *et al.*, 2001; Stapleton, 2003; George e Varmaz, 2005; Zhang *et al.*, 2006; Baggio-Zappia e Granato, 2009).

Amostras clínicas colhidas na era pré-HAART foram submetidas a um estudo (*Multicenter Acquired Immunodeficiency Syndrome Cohort Study*, desenvolvido nos EUA), na sequência do qual foi demonstrado que a virémia por GBV-C estava significativamente associada a um prolongamento na sobrevivência entre indivíduos do sexo masculino, HIV positivos, e que o desaparecimento de virémia de GBV-C em 5-6 anos após a seroconversão com HIV, estava relacionado com um pior prognóstico nestes doentes (Williams *et al.*, 2004). Estes indícios foram posteriormente confirmados já na era HAART (pós 1996) pela evidência de que os doentes co-infectados com replicação activa de GBV-C tinham melhor qualidade de vida (Tillmann *et al.*, 2004), contagens de células T CD4⁺ mais elevadas (Yeo *et al.*, 2000), e

melhores respostas iniciais à terapia anti-retroviral, com manutenção de supressão do vírus prolongada e menor risco de recaída viral (Rodriguez *et al.*, 2003; Mosam *et al.*, 2007).

Um particular interesse nos efeitos anti-retrovirais e imunológicos resultantes da terapia anti-retroviral em doentes infectados com GBV-C tem sido manifesto. Brumme e colaboradores mostraram que, apesar de uma elevada taxa virémia de GBV-C (20,4%) entre os doentes HIV seropositivos analisados, a presença de GBV-C, não surtiu qualquer efeito no início da resposta à terapia anti-retroviral (Brumme *et al.*, 2002). Por outro lado, relatórios posteriores, mostraram que os doentes co-infectados com GBV-C/HIV exibiam uma melhor resposta a HAART (Rodriguez *et al.*, 2003), nomeadamente entre os doentes que interromperam a terapia, onde se verificou uma diminuição da virémia de GBV-C inversamente proporcional à taxa de replicação do HIV, suportando a teoria de que a replicação de GBV-C pode estar relacionada com a dinâmica da infecção por HIV (Baggio-Zappia e Granato, 2009). Assim, a longa persistência da virémia de GBV-C é, provavelmente, uma componente chave para o efeito benéfico que resulta da co-infecção com GBV-C/HIV em doentes infectados, uma vez que a correlação inversa entre a carga viral de GBV-C e de HIV tem sido amplamente demonstrada (Lefrère^b *et al.*, 1999; Tillmann e Manns, 2001; Xiang *et al.*, 2001; Nunnari *et al.*, 2003; Williams *et al.*, 2004; Van der Bij *et al.*, 2005). Este facto pode ser uma explicação possível para as disparidades existentes nos resultados obtidos em diferentes estudos, onde o acompanhamento dos doentes tem sido por tempo mais reduzido e onde não foram observados efeitos positivos resultantes de uma situação de co-infecção GBV-C/HIV (Brumme *et al.*, 2002; Williams *et al.*, 2004; Bjorkman *et al.*, 2004; Van der Bij *et al.*, 2005). Para além disso, tem sido sugerido que o aumento da virémia com GBV-C na infecção com HIV poderá representar um fenómeno secundário no percurso natural da infecção pelo HIV, ao invés de um factor de prognóstico independente (Reshetnyak *et al.*, 2008).

Devido à heterogeneidade observada nas populações co-infectadas por HIV/GBV-C, surgem problemas a nível de metodologia quando se tenta comparar os resultados de diferentes estudos. A relevância destes casos é apontada quando não há acompanhamento da virémia do GBV-C longitudinalmente, sabendo-se, à partida, que a situação serológica do GBV-C pode alterar-se com o tempo, e a distinção entre os três estadios serológicos possíveis do GBV-C (presença de virémia ou de anticorpos anti-E2 ou ausência de ambos) é crucial para a interpretação dos resultados obtidos (Hoffmann *et al.*, 2007). Estas contradições podem ser justificadas, em parte, por diferenças a nível da população estudada, factores relacionados com o hospedeiro, a carga viral e introdução de terapia anti-retroviral (HAART, *Highly Active AntiRetroviral Treatment*),

distintos níveis de linfotropismo dos diferentes genótipos de GBV-C presentes e/ou os subtipos de HIV associados, os quais podem mascarar os putativos efeitos benéficos de GBV-C, contribuindo para resultados discrepantes encontrados nos diferentes grupos de pesquisa. (Abraham, 2007; Baggio-Zappia e Granato, 2009). Para além disso, a presença de uma co-infecção com HCV, pode influenciar no estudo da interacção HIV/GBV-C, uma vez que os vírus de hepatites partilham as mesmas vias de transmissão.

1.4.2.1. Mecanismos propostos para a interacção GBV-C/HIV

A identificação do(s) mecanismo(s) através do qual o GBV-C inibe a replicação do HIV pode(m) servir como base para o desenvolvimento de novas estratégias terapêuticas destinadas ao tratamento de doentes seropositivos para HIV (Baggio-Zappia e Granato, 2009). O mecanismo imunomodulador que parece ser responsável pelo efeito protector que, aparentemente, o GBV exerce numa co-infecção com o HIV, está ainda em estudo. No entanto, trabalhos recentes têm identificado um número de vias possíveis que permitem explicar a interacção entre ambos os vírus.

Desde que um estudo realizado por Heringlake e colaboradores demonstrou que, *in vitro*, a infecção com GBV-C pode tornar as células T CD4⁺ menos susceptíveis à infecção com HIV (Heringlake *et al.*, 1996), surgiu a hipótese de que esta conjugação pode estar relacionada com alterações na modulação dos receptores celulares do HIV através da produção de quimiocinas. Dados obtidos em 2005 permitiram verificar que a infecção de PMBCs com o GBV-C, levava a uma diminuição na replicação de isolados clínicos e laboratoriais de HIV, por decréscimo da expressão dos receptores de quimiocinas (CCR5 e CXCR4) à superfície das células T CD4⁺ (Jung *et al.*, 2005), uma vez é através destes receptores que o HIV adsorve às células susceptíveis (Figura 1.4). Anteriormente, foi ainda colocada como hipótese, o facto de o GBV-C poder ter a capacidade de induzir a síntese das quimiocinas RANTES (*Regulated on Activation Normal T cell-Expressed and Secreted*), MIP-1 α (*Macrophage Inflammatory Protein 1 α*), e *Macrophage Inflammatory Protein 1 β*) os quais são os ligandos naturais para o CCR5, assim como a proteína SDF-1 (*Stromal-Derived Factor*), sendo este o único ligando de CXCR4, originando assim, um decréscimo na expressão destes co-receptores (Xiang *et al.*, 2004).

Desta forma, a interacção entre as referidas quimiocinas com os seus receptores, promove uma internalização do receptor, razão pela qual ocorre um decréscimo da expressão dos co-receptores

que medeiam a entrada do HIV na célula alvo, e, conseqüentemente uma diminuição na replicação do vírus (Nattermann *et al.*, 2003; Peterlin e Trono, 2003).

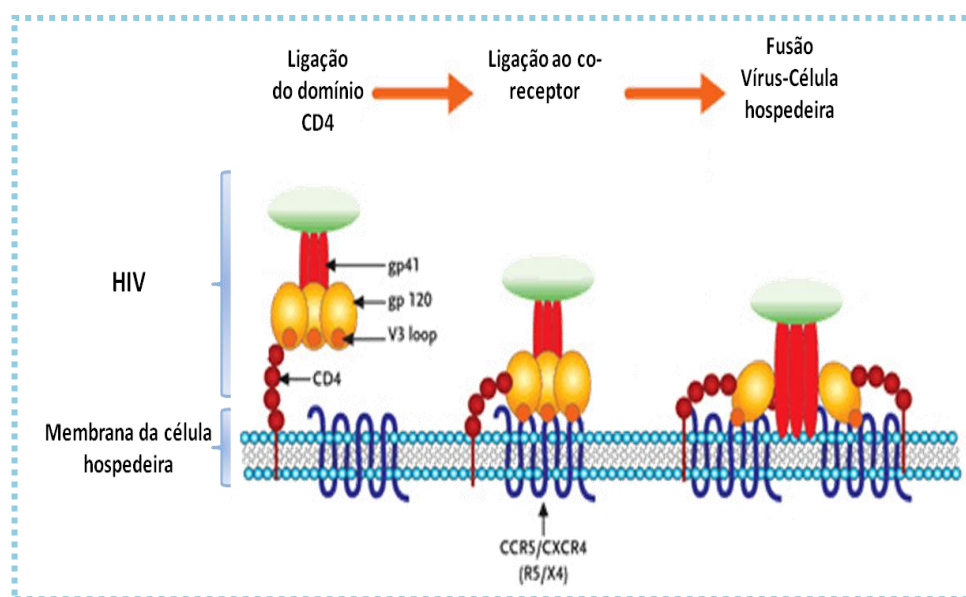


Figura 1.4. Esquema simplificado representativo da fusão do HIV à célula alvo através do receptor $CD4^+$, uma glicoproteína expressa, por exemplo, à superfície dos linfócitos T $CD4^+$, e dos co-receptores celulares CCR5 ou CXCR4. (Adaptado de Hoffmann *et al.*, 2007; Kopetzki *et al.*, 2008).

Na sequência destes trabalhos e de acordo com a hipótese inicialmente levantada, foi demonstrado que, PBMCs co-infectados com HIV e GBV-C expressam níveis mais altos de RNA mensageiro (mRNA) de RANTES, MIP-1 α , MIP-1 β e SDF-1, o qual está associado a um aumento da secreção destas quimiocinas *in vitro* quando comparadas com as células controle que estavam apenas mono-infectadas com HIV (Xiang *et al.*, 2004; Jung *et al.*, 2005). Particularmente, a ligação da glicoproteína E2 da superfície do GBV-C à proteína CD81 na superfície de linfócitos T, promove a libertação de RANTES (Nattermann *et al.*, 2003; Stapleton *et al.*, 2004; Xiang *et al.*, 2004; Kaufman *et al.*, 2007), com conseqüente diminuição da expressão celular de CXCR4. Ensaio realizados nos últimos 4 anos revelaram, ainda, que os anticorpos anti-E2 têm a capacidade de inibir directamente a replicação do HIV *in vitro*, através dos co-receptores CXCR4 e CXCR5 (Xiang *et al.*, 2006; Jung *et al.*, 2007). Recentemente, Mohr e Stapleton propuseram que a inibição da replicação do HIV pela proteína E2 seja dependente da carga viral de GBV-C (Mohr e Stapleton, 2009). Para além disso, o mesmo estudo identificou ainda um putativo péptido de fusão da proteína E2, que interfere com a oligomerização do péptido gp41 de HIV, necessário para a fusão do HIV-1, aludindo que a proteína E2 do GBV-C possa inibir a entrada do HIV na célula hospedeira (Herrera *et al.*, 2009; Mohr e Stapleton, 2009).

Analogamente, outras proteínas codificadas pelo GBV-C podem interferir, directa ou indirectamente, com a replicação do HIV. Xiang e colaboradores demonstraram que a proteína NS5A regula negativamente a replicação do HIV-1 em linfócitos *in vitro* (Xiang *et al.*, 2006). Os dados apresentados por estes autores revelaram que uma sequência de 85 resíduos de aminoácidos derivada da sequência primária de NS5A, é capaz de induzir a produção de RANTES e, desta forma, inibir a replicação do HIV (Xiang *et al.*, 2006). Estes resultados revestem-se de um particular interesse uma vez que abrem novas perspectivas ao desenvolvimento de uma nova classe de anti-retrovirais para o tratamento dos indivíduos infectados pelo HIV.

Estudos efectuados *in vivo*, descrevem perturbações complexas no perfil de citocinas em indivíduos co-infectados por GBV-C/HIV, sugerindo que o GBV-C exerce, ainda, um efeito imunomodulador na infecção com HIV, ao ajudar a manter níveis estáveis de citocinas *T-helper* 1 (Th1) durante o curso da infecção por HIV. Os doentes mono-infectados com HIV apresentam progressivamente níveis elevados de citocinas Th2 no soro (como a IL-4 e a IL-10), enquanto os doentes co-infectados com GBV-C têm um aumento na expressão de citocinas Th1, tais como IL-2 e IL-12 (Nunnari *et al.*, 2003). A importância da resposta das células *T-helper* (Th) contra o HIV é evidenciada pela correlação entre a progressão para estadio SIDA, com a incapacidade das PBMC produzirem IL2, IL12 e interferão- γ , e pelo aumento concomitante na secreção de IL-4 e IL-10 (Spellberg e Edwards, 2001; Breen, 2002). A IL-2 é uma citocina que interacciona com os linfócitos induzindo a proliferação celular de células T CD4⁺, razão pela qual, é utilizada no tratamento de infecções com HIV. No entanto, e por outro lado, sabe-se que a replicação de GBV-C é inibida, *in vitro*, por activação de culturas primárias de células CD4⁺ com IL-2 e fitohemaglutinina. Os resultados deste estudo mostram que a virémia de GBV-C está associada a um decréscimo na contagem das células CD4⁺ na presença de IL-2, pelo que a presença de GBV-C pode afectar a resposta dos doentes co-infectados ao tratamento com IL-2 (Mohr e Stapleton, 2009; Stapleton *et al.*, 2009).

Outro mecanismo proposto para a participação de GBV-C na imuno-modulação da replicação do HIV, é o da inibição da apoptose das células alvo, como resultado da detecção de baixos níveis de CD95 (o receptor da proteína Fas) à superfície dos linfócitos de indivíduos infectados por HIV com virémia GBV-C activa (Mönkemeyer *et al.*, 2008). Estes resultados sugerem que o GBV-C pode ter a capacidade de inibir a apoptose nos doentes co-infectados com HIV, e deste

modo, contribuir para maiores contagens de células $CD4^+$ encontradas nos indivíduos co-infectados com ambos os vírus.

A Figura 1.5 resume alguns dos mecanismos que têm sido apontados como possíveis explicações para o efeito benéfico que o GBV-C pode exercer contra o HIV numa situação de co-infecção. Contudo, permanecem por compreender muitos dos possíveis mecanismos de interação entre HIV e GBV-C, e os papéis que desempenham os factores do hospedeiro, específicos de cada indivíduo.

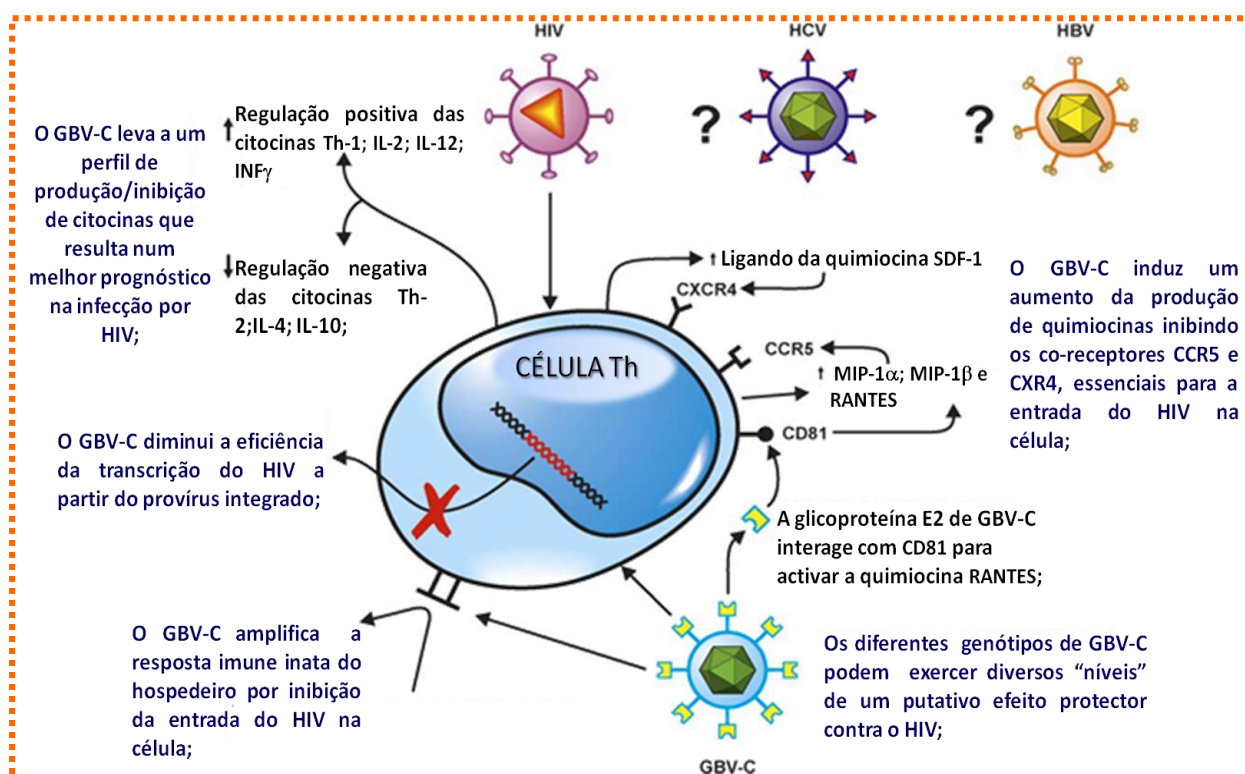


Figura 1.5. Múltiplos mecanismos propostos para a interferência do GBV-C no organismo hospedeiro numa situação de co-infecção com HIV (Adaptado de Stapleton *et al.*, 2004; Berzsényi^a *et al.*, 2005; George e Varmaz 2005; Baggio-Zappia e Granato 2009);

1.4.3. Co-infecção com HCV

Em situações de co-infecção com HCV e HIV, estudos mostram que, nestes indivíduos, existe uma evidência clara de que a doença hepática provocada por HCV é acelerada, com um aumento significativo na progressão da fibrose hepática (Benhamou *et al.*, 1999; Bowden, 2001; Rodríguez-Torres *et al.*, 2007). Nos trabalhos em que foram monitorizados indivíduos seropositivos para HIV sob terapêutica HAART, observou-se que esta exercia um efeito benéfico

no estado de saúde destes indivíduos numa situação de co-infecção com HCV (Voirin *et al.*, 2002; Qurishi *et al.*, 2003), embora a hepatotoxicidade associada aos fármacos utilizados em situações de hepatite viral crónica ou cirrose, seja uma preocupação constante (Berzsényi^a *et al.*, 2005). Uma vez que, só nos EUA, entre a população de UDIs (Utilizadores de Drogas Injectáveis), existe uma prevalência de co-infecção por HIV/HCV de 85% (Schooley, 2005), com base nos resultados obtidos pelos estudos realizados no âmbito das co-infecções GBV-C/HIV, tornou-se relevante ponderar se o GBV-C teria alguma interferência nos indivíduos triplamente infectados com HIV e HCV (Rendina *et al.*, 2001; Berzsényi^a *et al.*, 2005). Neste sentido, Sarabia e colaboradores evidenciaram uma influência benéfica resultante da co-infecção do vírus da hepatite C com GBV-C/HIV, na contagem de células CD4⁺ e na carga viral de HIV, por um período de 1 ano após interrupção de tratamento com HAART (Sarabia *et al.*, 2006). Por outro lado, foi sugerida uma associação da virémia de GBV-C a uma moderação significativa da severidade relacionada com a doença hepática originada por HCV verificando-se, por isso, uma redução na morbilidade provocada por doença hepática (Berzsényi^a *et al.*, 2005). Em paralelo, os doentes co-infectados com HIV/HCV com replicação activa de GBV-C, evidenciaram um menor grau de lesões histológicas e cargas virais de HCV, quando comparados com doentes apenas co-infectados com HIV/HCV. Até à data, pouco se sabe acerca dos factores relevantes para a subsistência ou eliminação da replicação do GBV-C.

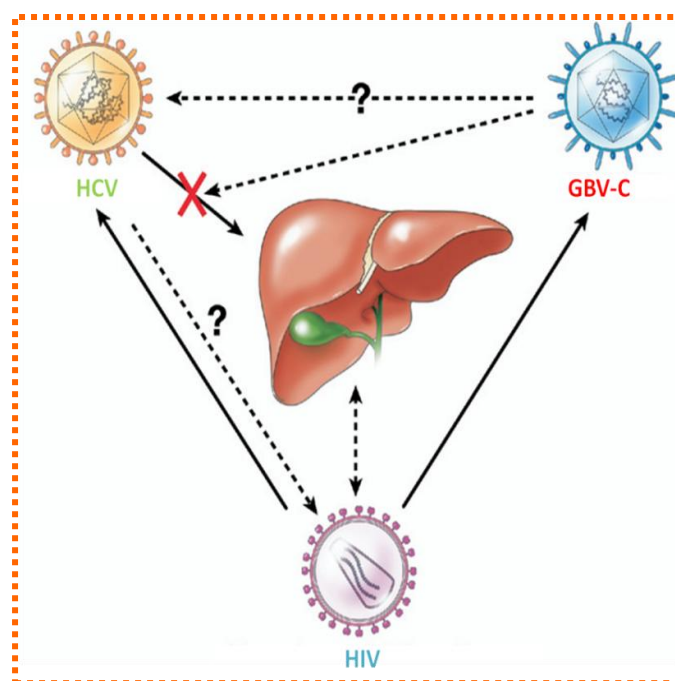


Figura 1.6. Representação esquemática de putativas interacções entre os vírus GBV-C, HCV, HIV e a doença hepática. (Adaptado de Stapleton, 2007).

Permanece sem resposta o facto da eliminação da virémia do GBV-C induzida pelo tratamento com interferão, terá o mesmo impacto no decurso da infecção por HIV e/ou HCV relativamente aos casos de eliminação espontânea observados. Esta questão é extremamente relevante a nível do aconselhamento na presença de co-infecções com HIV/GBV-C ou com HCV. Como tal, existe pelo menos a necessidade de rastreio do estado serológico relativamente ao GBV-C, do aconselhamento individual e do acompanhamento prospectivo durante a terapia de interferão em estudos controlados (Hoffmann *et al.*, 2007; Martini *et al.*, 2008).

1.5. Caracterização epidemiológica do vírus GBV-C

1.5.1. Distribuição geográfica e variabilidade genotípica

Desde que foi descoberto, têm sido vários os estudos que visaram avaliar a diversidade genética das diferentes estirpes de GBV-C globalmente distribuídas pelas populações humanas. De um modo genérico, e à semelhança do verificado para o HCV, as estirpes de GBV-C são classificadas em genótipos (Muerhoff *et al.*, 2006). No entanto, e ao contrário do registado para o HCV, o GBV-C é geneticamente mais conservado, exibindo apenas 11%-14% de divergência intergenotípica, comparativamente a um valor de 30% calculado para o HCV (Smith, 1997; Simmonds^b, 2001; Muerhoff *et al.*, 2005).

Os primeiros estudos de caracterização genética do GBV-C conduziram à descrição de 3 genótipos virais e apontavam para uma relação directa entre estes, e a origem étnica ou geográfica dos indivíduos infectados (Muerhoff *et al.*, 1996; Muerhoff *et al.*, 1997; Okamoto *et al.*, 1997). Neste grupo, estavam incluídos isolados do genótipo 1, tentativamente divididos em dois subtipos (1a e 1b) associados a indivíduos originários da África Ocidental, os subtipos 2a e 2b do genótipo 2, identificados em indivíduos norte-americanos e europeus (respectivamente), e o genótipo 3, o qual foi reportado em indivíduos japoneses (Muerhoff *et al.*, 1996; Muerhoff *et al.*, 1997; Muerhoff *et al.*, 2005). Mais tarde, um grupo japonês sugeriu ainda que dentro do genótipo 1 existissem pelo menos 5 subtipos putativos (Liu *et al.*, 2003). Análises filogenéticas subsequentes, de sequências nucleotídicas das regiões 5'UTR e NS5B evidenciaram a presença de estirpes do genótipo 2 de GBV-C na Rússia, Cazaquistão, Quirguistão, Turquemenistão, América do Sul, Paquistão, Japão e África Oriental (Sathat, 1999; Reshetnyak *et al.*, 2008). O genótipo 3, para além de estar presente no continente asiático, foi também reportado em populações indígenas da América do Sul (Naito e Abe, 2001; Loureiro *et al.*, 2002).

O número de genótipos de GBV-C expandiu-se quando foram estudadas novas estirpes virais pela análise da região 5'UTR, descobertas em Mianmar, no Vietname (Naito *et al.*, 1999 5'UTR), e na África do Sul (Sathar *et al.*, 1999), com ambos os grupos de investigação a reivindicar a identificação do quarto genótipo. No entanto, a análise da região 5'UTR e do gene E2 destas novas estirpes, em conjunto com representantes dos 3 genótipos previamente descritos, comprovaram a existência não de quatro, mas de cinco genótipos. Foi então proposto que as novas estirpes do Sudeste Asiático e Sul Africanas fossem incluídas, respectivamente, nos genótipos 4 e 5. (Smith *et al.*, 2000; Tucker e Smuts, 2000; Sathar e York 2001). Posteriormente, através de análises da região 5'UTR, foi ainda evidenciada a existência de um sexto genótipo, fundamentada na análise de um vírus identificado num indivíduo originário da Indonésia (Muerhoff *et al.*, 2006 86). A ilustração da distribuição global dos diferentes genótipos de GBV-C encontra-se representada na Figura 1.7.



Figura 1.7. Representação esquemática aproximada da distribuição geográfica do vírus GBV-C. Os diferentes genótipos virais encontram-se indicados numericamente (Adaptado de Polgreen *et al.*, 2003).

Na maioria dos trabalhos realizados, a identificação de novos genótipos recorreu a uma análise filogenética baseada no estudo da sequência da região 5'UTR em virtude do facto de esta ser a região do genoma com maior nível de conservação (Muerhoff *et al.*, 1996). Contrariamente aos demais flavivírus, e no que diz respeito ao HCV em particular, onde a genotipagem baseada na análise parcelar do genoma ou na sua sequência completa dá origem a resultados idênticos, o mesmo não acontece com o GBV-C (Simmonds^b, 2001). Por esta razão, as sequências das

regiões 5'UTR (Muerhoff *et al.*, 1997; Smith *et al.*, 1997) e E2, são as mais usadas para demonstrar, confirmar decisivamente a existência de diferentes genótipos e estabelecer uma correlação geográfica, uma vez que apenas estas, revelam relações filogenéticas congruentes com as obtidas quando é analisada a sequência genómica completa (Smith *et al.*, 2000).

1.5.1.1. Co-evolução dos vírus GBV-C com os seus hospedeiros humanos

Dada a distinta distribuição geográfica das variantes do GBV-C, é sugerido que esta esteja relacionada a uma longa co-evolução deste vírus com as populações humanas durante a sua migração global no período paleolítico (Smith *et al.*, 1997; Smith *et al.*, 2000; Naito e Abe 2001). Análises filogenéticas demonstram ainda que a diversidade registada nas estirpes africanas de GBV-C é maior do que a encontrada nos demais genótipos, colocando a possibilidade de estes vírus terem emergido a partir de África (Tanaka *et al.*, 1998; Sathar *et al.*, 1999; Muerhoff *et al.*, 2005), e como tal, suportando a teoria da co-evolução.

Por outro lado, a proximidade entre isolados de GBV-C provenientes do Sudeste asiático com variantes de origem no continente africano, parece confirmar os grandes fluxos migratórios de população humana de África para essa região durante o período Pleistocénico (Simmonds^b, 2001). Enquanto que as sequências das populações do Sudeste Asiático pertencem quase invariavelmente ao genótipo 3, este genótipo também é encontrado em populações nativas da América do Norte e do Sul (Tanaka *et al.*, 1998; Simmonds^b, 2001). Em contraste, as populações Caucasianas e outras originárias da Índia e do Norte de África, estão infectadas com o genótipo 2. O genótipo 1 está confinado à África sub-Sahariana, e evidencia, como referido anteriormente, o maior conjunto de diversidade a nível de sequências (Muerhoff *et al.*, 1997; Smith *et al.*, 1997; Liu *et al.*, 2000). Em todos os casos, estes são congruentes com as distribuições esperadas, caso o GBV-C já tivesse infectado o Homem no período de migração para fora do continente africano há aproximadamente 100.000-150.000 anos atrás (Tanaka *et al.*, 1998; Liu *et al.*, 2000; Mison *et al.*, 2000; Simmonds^b, 2001). Esta hipótese é ainda concordante com a baixa taxa de substituições nucleotídicas de GBV-C (estimadas em valores inferiores a $9,0 \times 10^{-6}$ por sítio, por ano (Suzuki *et al.*, 1999). O valor apontado, reflecte uma elevada contra-selecção relativa à ocorrência de substituições não sinónimas (d_N , locais onde as substituições nucleotídicas alteram o aminoácido codificado, dando origem a outro), 50 vezes inferior à variabilidade encontrada nos sítios sinónimos (d_S , não ocorre alteração de aminoácidos) (Simmonds e Smith, 1999), e evidenciada pelo alto grau de conservação da sequência poliproteica, cerca de 95%, mesmo entre

estirpes de áreas geográficas isoladas (Zuckerman, 1996; Naito e Abe, 2001). Embora a diversidade genética do GBV-C esteja associada, principalmente, à variação de sítios sinónimos (razão d_N/d_S significativamente inferior a 1), a ocorrência de pressões selectivas nas substituições silenciosas ficou demonstrada em regiões particulares do genoma (Simmonds^b, 2001). Tal facto advém, provavelmente, de limitações/restrições ao nível da função proteica, ou como consequência da presença de extensas estruturas secundárias no RNA viral (Simons *et al.*, 1996; Katayama *et al.*, 1998; Simmonds e Smith, 1999; Cuceanu *et al.*, 2001; Thurner *et al.*, 2004).

A evolução do GBV-C em humanos encontra-se, possivelmente, também espelhada noutros primatas (Charrel *et al.*, 1999). Existe uma grande divergência genética entre as estirpes isoladas em chimpanzés selvagens capturados na África Ocidental e Central, (Birkenmeyer *et al.*, 1998; Adams *et al.*, 1998) comparativamente à encontrada entre genótipos humanos. O facto de vírus relacionados com o GBV-C terem sido identificados em espécies de primatas do Novo Mundo sugere que os vírus tipo-GB podem estar amplamente distribuídos pelos símios, o que aponta para uma relação evolutiva próxima com os seus hospedeiros naturais (Adams *et al.*, 1998).

No entanto, ainda que a História evolutiva e diversidade genética global do GBV-C esteja relativamente estabelecida, a importância biológica da variabilidade genotípica tem sido questionada (Pavesi, 2001). De facto, alguns investigadores sugerem que o baixo nível de variabilidade observado nas sequências de aminoácidos entre os vírus GBV-C justifica a improvável existência de diferenças biológicas significativas entre estirpes classificadas em genótipos distintos (Simmonds e Smith 1999; Simmonds^b 2001).

1.5.2. Vias de transmissão

À semelhança do que acontece para o HBV, HCV e HIV, o GBV-C é facilmente transmitido por via parentérica (Figura 1.8). Esta pode ocorrer através de transfusões de sangue ou utilização dos seus derivados (ex. concentrados de factores de coagulação), partilha de agulhas e seringas utilizadas na injeção de drogas por via endovenosa (Berzsenyi^a *et al.*, 2005). No entanto, a prevalência das infecções por GBV-C na população em geral, com valores a variar entre 2 e 5%, é mais elevada do que a registada para outros vírus transmitidos pelas mesmas vias, sugerindo, como tal, a existência de outras vias de transmissão (Kelly e Skidmore, 2002). A transmissão vertical surge como uma das possíveis vias de disseminação do GBV-C (Reshetnyak *et al.*, 2008), podendo as infecções ocorrer durante a passagem do recém-nascido pelo canal do parto (Ohto *et al.*, 2000; Paternoster *et al.*, 2009). Existem, no entanto, registos de casos de crianças

infectadas com GBV-C pós-parto, sugerindo a aquisição de infecção comunitária, por exemplo no seio familiar, por vias de transmissão não identificadas (Pinho *et al.*, 1999; Lefrère^a *et al.*, 2000).

A detecção de GBV-C na saliva, sêmen e exsudados vaginais tem também sido registada (Quiros-Roldan, *et al.*, 2001; Bourlet *et al.*, 2002; Watanabe *et al.*, 2008), suportando, estas duas últimas situações, a possibilidade de transmissão viral por via sexual (Christensen *et al.*, 2003).

Embora o GBV-C possa ser transmitido tanto por contacto heterossexual como homossexual (Stark *et al.*, 1999; Reshetnyak *et al.*, 2008), as relações homossexuais entre indivíduos masculinos têm sido propostas como os veículos de transmissão mais eficazes deste vírus (Nerurkar, 1998; Berzsényi^b *et al.*, 2005).

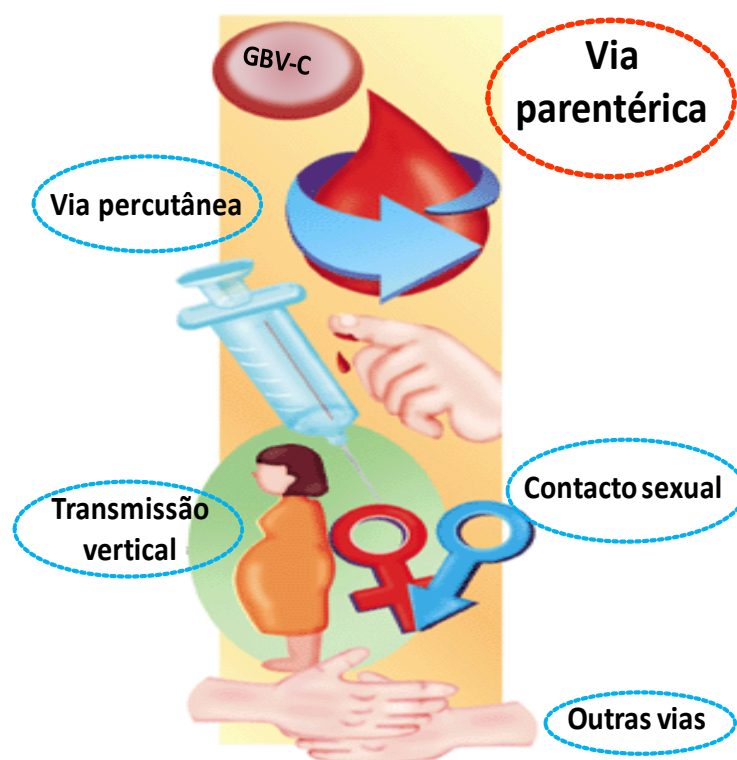


Figura 1.8. Principais vias de transmissão do vírus GBV-C (adaptado de http://www.info.gov.hk/hepatitis/leaflet/bil_02/indexe.htm).

1.5.3. Taxas de infecção pelo vírus GBV-C em diferentes grupos populacionais

As taxas de infecção pelo GBV-C variam globalmente entre 0,5 a 19% entre os dadores de sangue saudáveis (Figura 1.9) (El-Zayadi *et al.*, 1999; Alter *et al.*, 1997; Sathar *et al.*, 2000; Yan *et al.*, 2001; Stapleton, 2003; Mohr e Stapleton, 2009). Em aproximadamente 2% dos dadores de

sangue norte-americanos apresentam virémia e 17-20% apresentam anticorpos anti-E2 indicando um contacto prévio com o vírus (Feucht *et al.*, 1997; Sathar *et al.*, 2000).

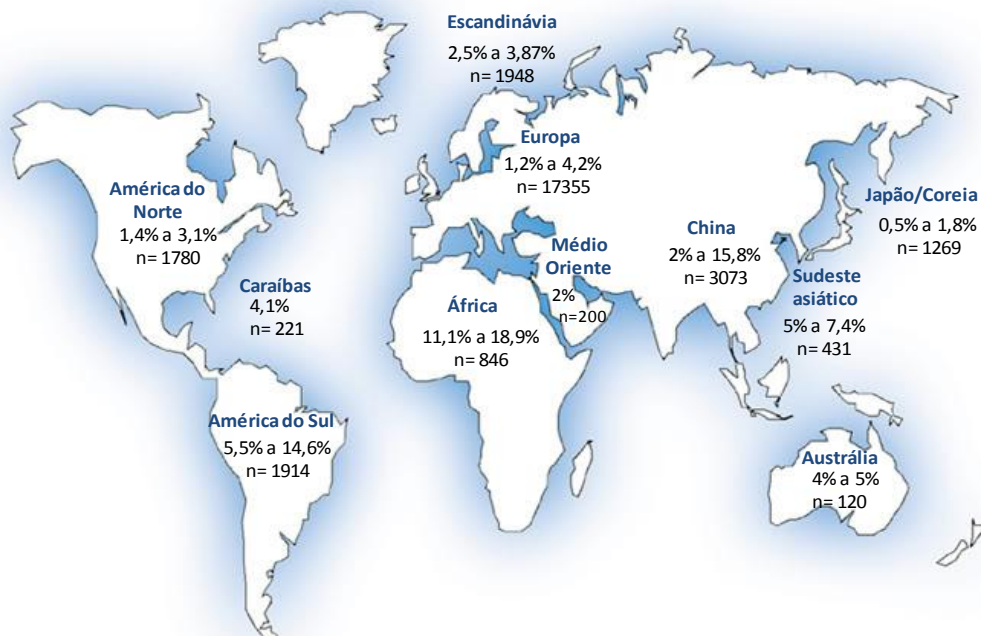


Figura 1.9. Representação das taxas de prevalência de GBV-C, entre os doadores de sangue, distribuídas globalmente (adaptado de Mohr e Stapleton, 2009).

A prevalência global de infecção pelo GBV-C varia de acordo com as características da população em estudo (Tabela 1.1). Contudo, é manifesta a presença comum de GBV-C associada a grupos de risco com aquisição da infecção por via percutânea e sexual (Berzsenyi^b *et al.*, 2005). A importância da transmissão via sexual é evidenciada pela elevada taxa de detecção de RNA de GBV-C em indivíduos homossexuais e prostitutas, em 63 e 25% dos casos, respectivamente (Stark *et al.*, 1999; Sawayama *et al.*, 1999).

Como esperado, os indivíduos sujeitos a hemodiálise apresentam uma prevalência de infecção pelo GBV-C mais elevada do que a população em geral (Feucht *et al.*, 1997; Hosseini-Moghaddam, 2008) podendo atingir percentagens na ordem dos 35,2% (Mauser-Bunschoten *et al.*, 1998), facto que está directamente relacionado com a duração do tratamento (Hayashi *et al.*, 1998) e com a co-infecção por HCV (Desassis *et al.*, 1999). No entanto, a percentagem de indivíduos que teve contacto com o vírus e que o conseguiu erradicar é bastante superior (51%) (Desassis *et al.*, 1999).

Tabela 1.1. Principais grupos de risco associados a infecção por GBV-C com indicação dos intervalos de amplitude encontrados em diversos estudos epidemiológicos

GRUPOS DE RISCO	PREVALÊNCIA	REFERÊNCIA(S) BIBLIOGRÁFICA(S)
UDIs	23% a 95%	Gutierrez <i>et al.</i> , 1997; Thomas <i>et al.</i> , 1998; Tillmann e Manns, 2001; Asim <i>et al.</i> , 2008.
Homossexuais masculinos	12% a 63%	Ibáñez <i>et al.</i> , 1998; Scallan <i>et al.</i> , 1998; Lefrère ^a <i>et al.</i> , 1999; Stark <i>et al.</i> , 1999; Rey <i>et al.</i> , 2000; Hattori <i>et al.</i> , 2003.
Prostitutas	14% a 25%	Scallan <i>et al.</i> , 1998; Sawayama <i>et al.</i> , 1999.
Indivíduos seropositivos para HIV	11% a 86%	Feucht <i>et al.</i> 1997; Xiang <i>et al.</i> , 2001; Stapleton, 2003; Williams <i>et al.</i> , 2004; Hekmat <i>et al.</i> , 2008.
Indivíduos seropositivos para HCV	10 a 74%	Masuko <i>et al.</i> , 1996; Alter <i>et al.</i> , 1997; Feucht <i>et al.</i> 1997; Yan e Dennin, 2000; Bjorkman <i>et al.</i> , 2001; Hoseini <i>et al.</i> , 2007; Asim <i>et al.</i> , 2008.
Co-infectados com HIV/HCV	30% a 85%	Schwarze-Zander <i>et al.</i> , 2006; Piroth <i>et al.</i> , 2008.
Indivíduos submetidos a múltiplas transfusões de sangue	5 a 38%	Jarvis <i>et al.</i> , 1996; Alter <i>et al.</i> , 1997; Feucht <i>et al.</i> , 1997; Hayashi <i>et al.</i> , 1998; Hosseini-Moghaddam <i>et al.</i> , 2008.

As prevalências associadas à transmissão vertical de GBV-C apresentam flutuações que podem situar-se entre os 50 e 90% (Wejstal *et al.*, 1999; Lefrère^a *et al.*, 2000; Kelly e Skidmore, 2002). Pelo facto de o GBV-C, o HCV e o HIV puderem ser transmitidos de forma idêntica, as co-infecções de GBV-C/HCV e/ou GBV-C/HIV são bastante comuns. Dados recentes do *Multicenter AIDS Cohort Study* (MACS) apontam para uma taxa de infecção com GBV-C entre os indivíduos seropositivos para HIV a rondar os 86% (Williams *et al.*, 2004), enquanto que evidências serológicas de infecção prévia com GBV-C encontram-se entre os 30% a 80% nos indivíduos seropositivos para HCV (Thomas *et al.*, 1998; Bjorkman *et al.*, 2001; Stapleton, 2003).

Por outro lado, os toxicodependentes utilizadores de drogas injectáveis (UDIs), são detentores das maiores percentagens de contacto com o GBV-C, com valores de taxa de infecção que podem atingir os 95% (Thomas *et al.*, 1998; Asim *et al.*, 2008).

Existe também uma grande variação na carga viral relativa ao GBV-C entre as diversas populações estudadas, o que poderá estar relacionado com diferenças a nível de infecciosidade.

1.5.4. Marcadores moleculares e detecção de GBV-C

As estimativas apontam para que 50 a 75% das infecções com GBV-C sejam resolvidas sem que seja detectada a replicação do vírus por métodos moleculares. Como referido em 1.3.4, são dois os marcadores serológicos da infecção por GBV-C. A presença de genomas virais circulantes (virémia), pode ser detectada usando métodos baseados na amplificação de secções do RNA viral por recurso à RT-PCR, enquanto que, a presença de anticorpos de GVB-C é revelada com recurso a ensaios imunoenzimáticos expressamente desenvolvidos para o efeito, onde os antígenos do invólucro, nomeadamente a glicoproteína E2, são utilizados como antígenos virais (Reshetnyak *et al.*, 2008). Como são mutuamente exclusivos, ou a virémia ou a presença de anti-E2 é detectável em indivíduos infectados com o GBV-C (Polgreen *et al.*, 2003).

A detecção laboratorial da infecção activa pelo GBV-C baseia-se, actualmente, em ensaios qualitativos que visam detectar a presença do RNA de GBV-C através de uma reacção de RT-PCR, usando como sequência alvo, a região 5'UTR por ser a mais conservada do genoma viral, conforme previamente mencionado em 1.5.1 (Bourlet *et al.*, 2002; Ruiz *et al.*, 2006; Souza *et al.*, 2006).

A hibridização *in situ* é reservada às tentativas de identificação dos putativos locais de replicação do GBV-C *in vivo* (Seipp *et al.*, 1999; Baggio-Zappia e Granato, 2009) enquanto que as reacções quantitativas de PCR em Tempo Real são usadas mais frequentemente na quantificação da carga viral (Castelain *et al.*, 2004; Jones *et al.*, 2005).

1.6. Notas finais/objectivos

Apesar das evidências experimentais e dos dados epidemiológicos parecerem apontar para que o hipotético papel do GBV-C no desenvolvimento de doença nos humanos (tal como inicialmente suposto) possa ser negligenciado, o estudo deste vírus tem suscitado um grande interesse na comunidade científica, pela associação com o putativo efeito protector que exerce contra a infecção pelo HIV. A literatura disponível, indica ainda que a infecção por GBV-C é persistente, e frequente, em grupos considerados de risco de aquisição de vírus por via parentérica, como o são os UDIs.

Considerando que a estimativa da prevalência e caracterização genética das estirpes de GBV-C circulantes em Portugal é bastante limitada, (Carvalho, *et al.*, 1999), não havendo quaisquer dados publicados na literatura internacional, tornou-se relevante aprofundar este conhecimento. Assim, constituíram objectivos deste trabalho:

- Determinar a taxa de infecção com GBV-C numa população maioritariamente constituída por utilizadores de drogas injectáveis e indivíduos seropositivos para HIV, residentes na área metropolitana da Grande Lisboa;
- Caracterizar a diversidade genética das estirpes virais encontradas, pela análise parcial de três regiões genómicas, 5'UTR/E1, E1/E2 e NS5A/NS5B. O fragmento 5'UTR/E1 inclui não só parte de uma região não codificante extremamente conservada (5'UTR), como também grande parte da região que codifica uma das glicoproteínas da superfície viral, designada E1. Por seu lado a análise de E1/E2 incidiu, em especial, sobre a porção do genoma viral que codifica a glicoproteína E2, a qual parece ser o principal alvo da resposta imunológica (humoral) do hospedeiro. Finalmente, o estudo da região NS5A/NS5B reflectiu-se, maioritariamente, na análise da secção genómica codificante da proteína não estrutural NS5B, a qual codifica a polimerase de RNA dependente de RNA.
- Entre outras abordagens, genotipar as variantes circulantes através de métodos de análise filogenética e identificar de polimorfismos genéticos nas estirpes virais estudadas pela análise das respectivas sequências de aminoácidos;

- Implementar ensaios de PCR em Tempo Real para estimar, de forma semi-quantitativa, a carga viral de GBV-C utilizando dois métodos de detecção distintos: ensaios com sondas *TaqMan* e ensaios com o corante *SYBR Green I*.

2. MATERIAL E MÉTODOS



MATERIAL E MÉTODOS

2.1. Caracterização da população estudada

O presente estudo envolveu a análise de amostras de sangue colhidas entre 1998 e 2009. Especificamente, 126 amostras foram colhidas, entre 1998 e 2001, no Hospital Fernando da Fonseca (Amadora Sintra), 2 amostras foram colhidas durante o ano de 2000 no Centro de Acolhimento a Toxicodependentes do Casal Ventoso e 87 amostras provieram do Centro de Atendimento a Toxicodependentes das Taipas, em Lisboa. De entre estas últimas, 8 foram colhidas entre 1999-2000 e as restantes entre 2008 e 2009. As colheitas de sangue, efectuadas mediante consentimento informado, foram acompanhadas do preenchimento de um inquérito epidemiológico que permitiu a obtenção de informação sobre uma série de parâmetros epidemiológicos básicos tais como a idade, género, nacionalidade, e comportamentos de risco (partilha de seringas para injeção de drogas por via endovenosa e relações sexuais não protegidas) dos indivíduos envolvidos neste estudo.

A informação referente ao estatuto serológico para HIV-1, HIV-2 e HCV foi fornecida por estudos realizados em paralelo. Estes dados estão sumariamente indicados na Figura 2.1.

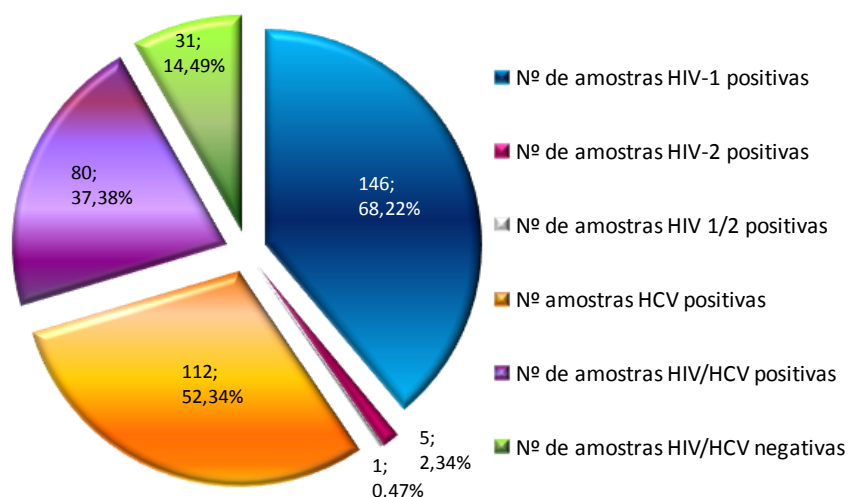


Figura 2.1. Caracterização da população estudada. Em cada uma das fatias do gráfico são indicados os números absolutos de amostras analisadas, assim como respectiva percentagem em relação ao total da população estudada.

Do total das 214 amostras analisadas, 152 eram provenientes de indivíduos seropositivos para HIV (39,47% dos quais reportaram terem sido infectados por contacto sexual, enquanto que

58,55% terão sido infectados através da partilha de material de injeção e em 1,98% dos casos, não foi reportada a via de infecção) e 112 amostras corresponderam a indivíduos seropositivos para HCV, em 80 dos quais, foi confirmada co-infecção com ambos os vírus HIV-1 e HCV. Finalmente, 31 dos indivíduos em estudo não revelaram seropositividade para qualquer dos vírus HIV ou HCV (Figura 2.1).

No total, a população estudada incluiu 151 indivíduos toxicodependentes.

Este estudo envolveu 135 indivíduos do sexo masculino (63,08%) com idades compreendidas entre os 21 e os 67 anos (média de idades=37 anos), e 79 do sexo feminino (36,92%) com idades compreendidas entre os 17 e os 64 anos (média de idades=32 anos). Destaca-se ainda o facto de o grupo ter incluído 67 indivíduos com nacionalidade não portuguesa, na sua maioria com etnia africana (17 Angolanos, 22 Cabo Verdianos, 12 Guineenses, 5 Moçambicanos, 4 São Tomenses e 1 Sul-Africano) (Figura 2.2). Indivíduos de nacionalidade Brasileira (n=3), Espanhola (n=1) e Francesa (n=2) foram também incluídos nesta amostragem.

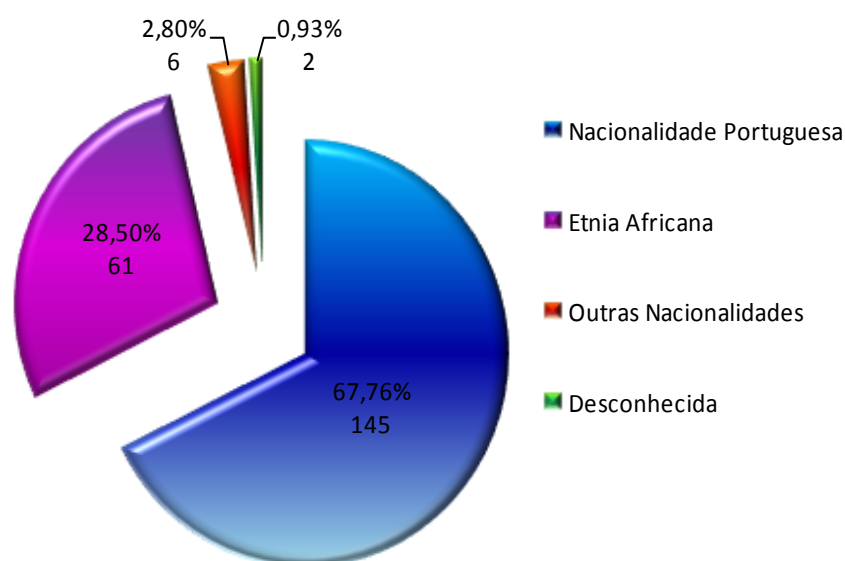


Figura 2.2. Caracterização da população estudada no que respeita à nacionalidade dos indivíduos envolvidos.

2.1.1. Produto biológico: colheita e preparação da amostra

Para o presente estudo foram seleccionadas 214 amostras de plasma sanguíneo humano, previamente separado por centrifugação (25000g durante 10min) a partir de 5-10ml de sangue periférico total, colhidos por punção venosa para tubos contendo 1ml de citrato de sódio a 3,8% (p/v) (agente anti-coagulante). As amostras de sangue foram mantidas à temperatura ambiente

(2-3 horas) até ao seu processamento. O plasma daí resultante foi distribuído em alíquotas devidamente identificadas, as quais foram posteriormente congeladas à temperatura de -80°C.

2.2. Extracção de RNA total a partir de plasma humano

A extracção de RNA total a partir de plasma humano, foi efectuada através da utilização do sistema comercial *INSTANT Virus RNA Kit* (AnalytikJena, Alemanha). Este sistema inclui tubos de extracção revestidos por um transportador (*carrier*) de ácidos nucleicos, bem como um controlo interno de extracção, o qual permite facilitar a interpretação de falsos resultados negativos nas reacções de amplificação subsequentes, consequência, por exemplo, da presença de inibidores nos extractos de RNA. O procedimento foi realizado de acordo com as instruções do fabricante.

Genericamente, o protocolo de extracção pressupõe o descongelamento das amostras à temperatura ambiente, seguido de um passo de lise em tubos de extracção, lavagem das colunas onde se fixou o RNA e, por fim, a eluição dos ácidos nucleicos extraídos. A etapa de lise foi efectuada em laboratório de segurança biológica BSL-2 (BSL-*BioSafety Level*), numa câmara de fluxo laminar, onde foram adicionados 450µl de uma solução de lise (RL) a cada tubo de extracção. Posteriormente, a esse volume adicionou-se 150µl de plasma (manipulação efectuada em laboratório de nível de segurança biológica BSL-3). As misturas plasma/solução de lise foram homogeneizadas vigorosamente 3 a 4 vezes num agitador de tipo vórtex, em pulsos de 10seg, durante um período de incubação de 15min à temperatura ambiente, seguida de centrifugação rápida. A extracção prosseguiu com a adição, às misturas anteriormente preparadas, de um volume de 600µl de solução tampão de ligação (RBS). Estas foram homogeneizadas por agitação forte (vórtex) e em seguida aplicadas nas colunas de extracção, previamente colocadas em tubos colectores. O processo foi iniciado com a adição às colunas de 650µl das misturas contendo o RNA, o qual a elas foi fixado durante uma breve centrifugação a 10000g durante 1min. O volume residual de cada amostra foi, então, adicionado à mesma coluna e o processo repetido. As lavagens das colunas, para remoção de contaminantes, foram efectuadas por adição sucessiva das soluções de lavagem HS seguida de LS, e pela centrifugação dos tubos de extracção a 10000g durante 1min. O RNA foi eluído por adição de 30µl de água livre de RNase a cada coluna, seguida de uma breve incubação de 2min à temperatura ambiente. O volume eluído foi recolhido, por centrifugação (1min a 10000g), para tubos limpos. O volume final de cada amostra foi dividido em duas alíquotas de 5µl e uma de 20µl. Foi repetida a etapa

de eluição uma segunda vez, de forma a obter uma alíquota extra de 30µl (recolha do RNA residual). Todas as alíquotas foram congeladas à temperatura de -80°C.

2.3. Síntese, por transcrição reversa (RT), de DNA complementar (cDNA) a partir de RNA total

As reacções de RT (*Reverse Transcription*) foram executadas usando os sistemas comerciais *Improm-IITM Reverse Transcription System* (Promega, EUA) e *RevertAidTM H Minus First Strand cDNA Synthesis Kit* (Fermentas, Lituânia). Em ambos os sistemas o procedimento foi adaptado relativamente às instruções sugeridas pelo fabricante. Para a síntese de cDNA foram utilizados 5µl do extracto de RNA total (preparado de acordo com o descrito em 2.2.), tendo sido utilizado um volume final de reacção de 20µl. A preparação de cada reacção foi efectuada no gelo.

O recurso ao sistema *Improm-IITM Reverse Transcription System* implicou a preparação de reacções de RT iniciadas pela adição de 1µl de solução contendo hexâmeros aleatórios (oligonucleótidos, ou *primers*, de sequência aleatória), a cada alíquota de 5µl do extracto de RNA total. Após agitação manual suave (inversão) e breve centrifugação, a mistura anterior, juntamente com o tubo correspondente ao controlo negativo, preparado com 5µl de água livre de nucleases e 1µl dos mesmos hexâmeros, foram mantidos a 70°C durante 5min, de forma a desnaturar as moléculas de RNA (remoção de estruturas secundárias). Os tubos foram imediatamente colocados num banho de gelo fundido por igual período de tempo. As misturas de reacção foram preparadas numa câmara de fluxo laminar. Estas incluíram 5,1µl de H₂O livre de nucleases, 4µl de solução tampão de reacção (*ImpromIITM 5X*), 2,4µl de MgCl₂ (6mM final), 1µl de dNTPs (Desoxiribonucleótidos 5'-trifosfatados) (concentração final de cada dNTP, igual a 0,5mM), 0,5µl de inibidor de ribonucleases (*Recombinant RNasin*) e 1µl de transcriptase reversa (*Improm-IITM Reverse Transcriptase*), perfazendo um volume total de 14µl. Após a homogeneização das misturas de reacção, estas foram adicionadas de 6µl das misturas RNA/hexâmeros. A mistura resultante foi colocada num termociclador *myCycler* (Bio-Rad, EUA) programado como descrito na Tabela 2.1. O produto final (cDNA) foi mantido a 4°C para subsequente (no próprio dia) amplificação de sequências genómicas virais através da reacção de polimerização em cadeia da polimerase (PCR, *Polymerase Chain Reaction*), ou conservado a -20°C (para utilização posterior).

Tabela 2.1. Condições de reacção utilizadas nas experiências de síntese de cDNA.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C)	TEMPO DE DURAÇÃO
Hibridação	1	25°C	5'
Extensão	1	42°C	60'
Inactivação da enzima	1	70°C	15'

Quando, em alternativa, foi utilizado o sistema *RevertAidTM H Minus First Strand cDNA Synthesis*, o método usado foi semelhante ao descrito anteriormente. As misturas de RT foram preparadas por adição de 4µl de solução tampão de reacção concentrada 5X, 1µl de inibidor de RNases (*RiboLockTM RNase Inhibitor*), 2µl de uma mistura de dNTPs concentrados a 10mM (concentração final 1mM) e 1µl de transcriptase reversa (*RevertAidTM H Minus M-MuLV*). Foi também preparada uma pré-mistura com 6µl de água tratada com dietil-pirocarbonato (DEPC) juntamente com 1µl de hexâmeros aleatórios para cada reacção de transcrição reversa a realizar. A esta mistura foram adicionados 5µl de extracto de RNA total, enquanto que ao tubo correspondente ao controlo negativo da reacção de RT foram adicionados 5µl de água tratada com DEPC (um inibidor não específico de RNases). As misturas de reacção foram, seguidamente, submetidas a aquecimento/arrefecimento rápido, como anteriormente descrito. Após este tratamento, as misturas RNA/hexâmeros foram adicionadas de 8µl da mistura de reacção, agitadas brevemente, e colocadas num termociclador *myCycler* (Bio-Rad, EUA). A reacção de RT decorreu nas condições descritas na Tabela 2.1.

2.4. Delineamento de oligonucleótidos sintéticos (*primers*) iniciadores de PCR

O desenho dos *primers* para amplificação, por *nested-PCR*, da região não traduzida localizada na extremidade 5' do genoma do GBV-C (5'UTR) e de parte das sequências nucleotídicas que codificam as glicoproteínas estruturais de superfície (E1 e E2), bem como as proteínas não estruturais NS5A e NS5B (polimerase de RNA dependente de RNA), foi elaborado manualmente, com recurso a um alinhamento múltiplo de sequências nucleotídicas, previamente construído com base em 44 sequências de referência disponíveis no GenBank (acedido através do portal do *NCBI, National Center for Biotechnology Information* em <http://www.ncbi.nlm.nih.gov/>), as quais contemplaram os 6 génotipos reconhecidos de GBV-C.

O alinhamento das sequências em questão foi efectuado com o programa informático *MAFFT version 6* (Kato *et al.*, 2008; disponível em <http://align.bmr.kyushu-u.ac.jp/mafft/online/server>), e baseou-se na pesquisa de regiões conservadas entre as várias sequências nucleotídicas alinhadas, a montante e a jusante das diferentes regiões a analisar. A Tabela 2.2 inclui a designação, sequência, tamanho, temperatura de hibridação e coordenadas relativas a uma sequência de GBV-C de referência (NC_001710) para cada um dos *primers* utilizados neste trabalho.

Tabela 2.2. Informação relativa ao conjunto de *primers* delineados para amplificação de fragmentos de DNA de GBV-C por *nested*-PCR.

REGIÃO DO GENOMA	PRIMER (DESIGNAÇÃO)	SEQUÊNCIA (5'→3')	TAMANHO (NT)	TM (°C) ¹	COORDENADAS (NC_001710)
5'UTR + 5'UTR/E1	UTRForO	TGGTAGCCACTATAGGTGGGTCTTAAG	27	60°C	160 – 186
	UTRForI	GCGCACGGTCCACAGGTGTTGGCCCTAC	28	69°C	224 – 251
	E1rev1	GGRGCTGGGTGGCCYCATGCWTTTCCC	27	66°-69°C	450 – 476
	E1rev2	AAATGCCACCCGCCCTACCCGAAGG	26	66°C	509 – 534
5'UTR/E1	E1rev3	GGGGCCACYGATTTTGYCCGTGGCYCC	27	66°-70°C	1375 – 1400
	E1rev4	RCTGKTCAGTCCATCTCCAAAACCTCAC	27	58°-61°C	967 – 994
E1/E2	E2For1	AGTGAGTTTTGGAGATGGACTGARCAG	27	58°-60°C	966 – 993
	E2For2	GTGTGAYTGCCCCAAYGGYCCCTGGGT	27	64°-69°C	1295 – 1321
	E2rev2	CCACARCACRAGAACATBAGGCGYTG	27	58°-66°C	2433 – 2459
	E2rev3	GGGAAWGCYCCCCGAGCRAGCTTCCAC	27	66°-69°C	2456 – 2483
NS5A/ NS5B	NS5BForO	GCTGAYGTGGCTAGCYTGTGTGARATG	27	60°-64°C	7233 – 7259
	NS5BForI	ATCCAGAACCATACAGCCTATTGTGAC	27	58°C	7263 – 7289
	NS5Brev5	CCAAGTGGAGCAGAAGGGRGCYGTGTC	27	64°-67°C	8421 – 8447
	NS5Brev6	TGATGACCCACCGTGTGATGGGRTGCC	27	64°-66°C	8581 – 8607

Nota: posições degeneradas: W=A ou T; Y=T ou C; R=G ou A; B=C ou G ou T;

¹A temperatura de desnaturação (*melting*) dos *primers*, foi calculada com recurso à ferramenta *Oligo Calc: Oligonucleotide Properties Calculator*, disponível em <http://www.basic.northwestern.edu/biotools/oligocalc.html>. NT corresponde a “nucleótidos”, enquanto que TM indica a temperatura de desnaturação (do inglês *temperature of melting*).

Os *primers* foram desenhados de forma a evitar a existência de regiões de complementaridade na sequência de cada um deles, assim como entre pares de *primers* utilizados nas reacções de PCR. Onze dos 15 *primers* delineados apresentam degenerações em algumas posições de forma a acomodar a variabilidade genética entre as sequências alinhadas. O esquema representativo da posição relativa de cada um dos *primers* desenhados para amplificação das diferentes regiões em estudo, assim como o tamanho das sequências esperadas em cada reacção de PCR, estão indicados esquematicamente na Figura 2.3.

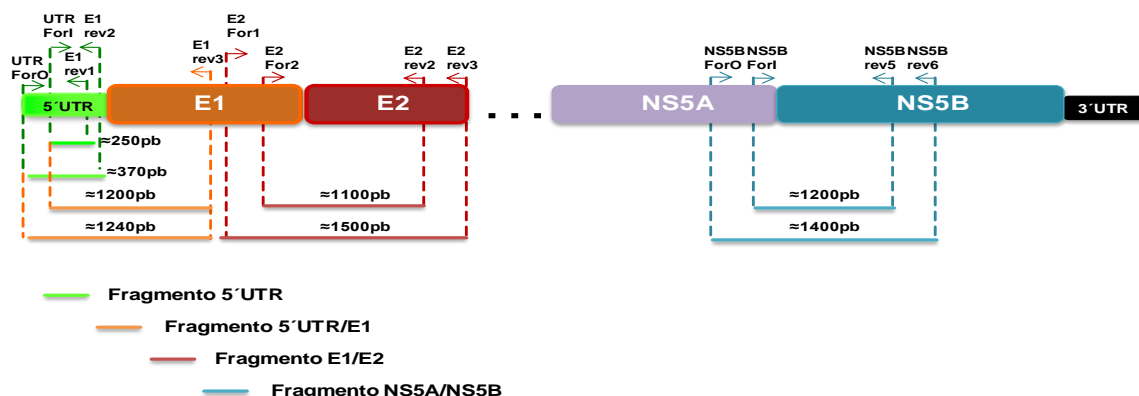


Figura 2.3. Esquema representativo da localização aproximada dos *primers* utilizados no decurso deste trabalho. O tamanho aproximado do produto de amplificação esperado para cada região estudada encontra-se também indicado. Por questões gráficas, a secção do genoma viral entre a região codificante da proteína NS2 e NS4B foi suprimida (indicado por ...).

2.5. Amplificação de sequências genómicas do vírus GBV-C por *nested*-PCR

2.5.1. Detecção da presença do vírus GBV-C: amplificação da região conservada 5'UTR

A detecção da presença do genoma do vírus GBV-C foi realizada através da amplificação, por *nested*-PCR, de uma das regiões mais conservadas do genoma viral, localizada no extremo 5' do genoma viral e designada por região UTR (5'UTR).

A técnica de *nested*-PCR envolve uma dupla reacção de PCR, e tem como objectivo aumentar a sensibilidade, assim como a especificidade, da amplificação pretendida. Este método pressupõe duas reacções de amplificação sucessivas, sendo que o produto resultante da primeira serve de matriz para a segunda reacção de amplificação. Na primeira reacção são utilizados *primers* ditos

externos, que se ligam a uma região externa àquela que se pretende analisar. O segundo par de *primers* (internos) liga-se ao produto da primeira reacção de PCR, numa região interna relativamente ao local de ligação dos *primers* externos, originando em consequência do processo de amplificação, um fragmento de tamanho inferior.

Para amplificação da região 5'UTR do genoma de GBV-C foram utilizados UTRForO e E1rev2 como *primers* externos e o par UTRForI e E1rev1 como *primers* internos (Tabela 2.2.). Em cada uma das reacções de amplificação usou-se o sistema *puReTaq Ready-To-Go PCR Beads* (GE Healthcare, Reino Unido). Neste sistema os reagentes utilizados para as reacções de PCR encontram-se liofilizados sob a forma de pequenas esferas, em concentrações optimizadas para um volume final de reacção de amplificação de 25µl. Cada esfera, contida em tubos de PCR individuais, contém estabilizadores de temperatura, BSA (Albumina sérica bovina), dNTPs (para uma concentração final de 200µM para cada um deles), MgCl₂ (1,5mM final), Tris-HCl (pH 9,0, 10mM final), KCl (50mM final) e 2,5U de uma polimerase de DNA recombinante termorresistente (*puReTaq DNA Polymerase*).

A preparação da mistura de reacção foi efectuada numa câmara de fluxo laminar. Esta preparação consistiu na hidratação de cada esfera utilizada com 16µl e 19µl de H₂O livre de nucleases na primeira e segunda reacção de PCR, respectivamente. A esse volume adicionou-se 2µl de cada uma das soluções contendo 10pmol de cada um dos *primers* a utilizar (800nM final em cada uma das reacções de amplificação). À mistura de reacção foi ainda adicionado DNA matriz contido em 5µl das misturas de RT preparadas como descrito em 2.3. (primeira reacção de PCR), e 2µl do produto da primeira reacção de PCR (segunda reacção de amplificação). As reacções de amplificação foram realizadas nos termocicladores *myCycler* (Bio-Rad, EUA) ou *iCycler* (Bio-Rad, EUA). As condições de amplificação foram comuns a ambas as etapas de PCR e encontram-se descritas na Tabela 2.3.

Tabela 2.3. Condições de reacção utilizadas na amplificação da região 5'UTR de GBV-C por *nested*-PCR.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C)	TEMPO DE DURAÇÃO
Desnaturação inicial	1	95°C	2'
Desnaturação	35	95°C	45''
Hibridação		55°C	45''
Extensão		72°C	45''
Extensão final	1	72°C	5'
Manutenção	1	4°C	∞

2.5.2. Amplificação da região 5'UTR/E1 do genoma do GBV-C

Para a amplificação de um fragmento do genoma viral que inclui parte da 5'UTR e a região que codifica o extremo N-terminal da glicoproteína E1 de superfície (Figura 2.3.), foi utilizado um protocolo de *heminested*-PCR. O princípio desta técnica é o mesmo do descrito para *nested*-PCR (ver secção 2.5.1.) com a diferença de que um dos *primers* utilizados é comum a ambas as reacções de amplificação. Na primeira reacção de PCR foi utilizado o sistema *puReTaq Ready-To-Go PCR Beads* (GE Healthcare, Reino Unido), utilizando as condições descritas em 2.5.1. e o par de *primers* externos UTRForO e E1rev3 (Tabela 2.2.). Após preparação das misturas de reacção, estas foram submetidas a 35 ciclos de amplificação, como indicado na Tabela 2.4., utilizando o termociclador *myCycler* (Bio-Rad, EUA).

Tabela 2.4. Condições de reacção utilizadas na primeira e segunda reacções de PCR destinada à amplificação da região 5'UTR/E1.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C) Primeira reacção de PCR	TEMPERATURA (°C) Segunda reacção de PCR	TEMPO DE DURAÇÃO
Desnaturação inicial	1	95°C	95°C	2'
Desnaturação	25	95°C	95°C	45''
Hibridação		55°C	58°C	45''
Extensão		72°C	72°C	2.30'
Desnaturação	10	95°C	95°C	45''
Hibridação		55°C	60°C	45''
Extensão		72°C	72°C	2.30' + 5'' *
Extensão final	1	72°C	72°C	5'
Manutenção	1	4°C	4°C	∞

*Incremento no tempo de extensão em 5'' com o objectivo de diminuir a perda de actividade da polimerase (programas de amplificação longos).

As segundas reacções de PCR foram efectuadas com dois sistemas de amplificação distintos. Numa primeira situação foi utilizada uma mistura de reacção pronta a usar (*2X PCR Master MIX* – Fermentas, Lituânia) contendo 0,05 U/μl *Taq DNA Polymerase* em solução tampão de reacção, MgCl₂ (2mM final) e dNTPs para uma concentração final de 200mM (cada um). Cada reacção de amplificação foi efectuada num volume final de 25μl aos quais foram adicionados 12,5μl de *2X PCR Master MIX*, 6,5μl de água tratada com DEPC, 2μl dos *primers* (800mM final)

correspondentes à segunda reacção de amplificação (UTRForI e E1rev3) e 2µl do produto de DNA obtido na primeira reacção de PCR.

Para os casos em que o rendimento do produto de PCR amplificado com este sistema se apresentou baixo recorreu-se, em alternativa, ao sistema *puReTaq Ready-To-Go PCR Beads*, tal como descrito em 2.5.1., tendo sido utilizados 2µl da primeira reacção de PCR como matriz para a segunda amplificação, e tendo esta decorrido segundo as condições descritas na Tabela 2.4.

2.5.3. Amplificação das regiões E1/E2 e NS5A/NS5B do genoma do GBV-C

As sequências de DNA que codificam as glicoproteínas de superfície E1 e E2 (região E1/E2) e as proteínas não estruturais NS5A e NS5B (NS5A/NS5B), foram amplificadas (parcialmente) por recurso à técnica de *nested*-PCR.

O procedimento realizado foi executado com os sistemas de amplificação e condições idênticas às aquelas descritas para a amplificação da região 5'UTR/E1, desta feita, mediante a utilização dos pares de *primers* E2For1/E2rev3 e E2For2/E2rev2 como *primers* externos e internos, respectivamente, na amplificação da região E1/E2, e os pares de *primers* NS5BForO/NS5Brev6 (internos) e NS5BForI/NS5Brev5 (externos) para a região NS5A/NS5B. As condições de reacção utilizadas para a amplificação de ambas as regiões foram comuns, encontrando-se descritas na Tabela 2.5.

Tabela 2.5. Condições de reacção utilizadas nas reacções de PCR destinadas à amplificação das regiões E1/E2 e NS5A/NS5B.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C)	TEMPO DE DURAÇÃO
Desnaturação inicial	1	95°C	2´
Desnaturação	25	95°C	45´´
Hibridação		55°C	45´´
Extensão		72°C	2.30´
Desnaturação	10	95°C	45´´
Hibridação		58°C	45´´
Extensão		72°C	2.30´ + 5´´ *
Extensão final	1	72°C	5´
Manutenção	1	4°C	∞

*Incremento no tempo de extensão em 5´´ com o objectivo de diminuir a perda de actividade da polimerase.

2.6. Análise de moléculas de DNA por electroforese em gel de agarose

Os produtos resultantes das reacções de amplificação executadas, foram detectados através da sua resolução por electroforese em gel de agarose, na presença de 0,5µg/ml de brometo de etídeo, tendo-se utilizado solução tampão TAE (40mM Tris-Acetato, 1mM EDTA, pH 8,0) como tampão de electroforese. Os fragmentos de menor massa molecular, resultantes da amplificação da região 5'UTR, foram separados em gel de agarose a 1,5% enquanto que os produtos de amplificação com uma massa molecular superior a 1kb (regiões 5'UTR/E1, E1/E2 e NS5A/NS5B) foram separados em géis de agarose a 1% (Anexo IV). Em todas as situações foram analisadas amostras de 3µl de cada um dos produtos que resultou das reacções de PCR. A massa dos fragmentos obtidos foi estimada através da análise comparativa da migração destes em gel de agarose, com a de um marcador de fragmentos de massas moleculares conhecidas (*GeneRulerTM 100bp DNA Ladder Plus-MBI* - Fermentas, Lituânia).

A visualização dos fragmentos de DNA foi conseguida através da exposição dos géis de agarose a radiação UV, e a imagem de cada um, captada digitalmente num aparelho *Gel-Doc XR* (BioRad, EUA).

2.7. Purificação de produtos de PCR

Os produtos de PCR desejados foram purificados de forma a puderem ser, de seguida, sequenciados. Para tal, utilizaram-se, em alternativa, dois sistemas comerciais: o *QIAquick PCR Purification Kit* (QIAGEN, EUA) e o *Wizard®PCR Preps DNA Purification System* (Promega, EUA). Em ambos os casos o procedimento experimental foi executado de acordo com as instruções do fabricante. O esquema representativo dos protocolos realizados é o exemplificado esquematicamente na Figura 2.4. Resumidamente, a utilização do sistema *QIAquick PCR Purification Kit* pressupôs a diluição, com água tratada com DEPC, do produto de PCR para um volume final de 100µl. De seguida, adicionou-se uma solução tampão de ligação (*PBI*), tendo as misturas assim obtidas sido transferidas para uma coluna (*QIAquick Spin Column*) onde ocorreu a fixação do DNA. Posteriormente, submeteu-se a coluna a uma lavagem com uma solução tampão adequada (*PE*), rejeitando o sobrenadante em cada uma das duas centrifugações consecutivas. O DNA foi eluído num volume de 30µl com solução tampão *EB*.

Por outro lado, o *Wizard® PCR Preps DNA Purification System* (Promega, EUA), permitiu a purificação de DNA através da sua passagem por minicolunas (*Wizard Minicolumns*), após a sua fixação prévia a uma resina catiónica (*Wizard® PCR Preps DNA Purification Resin*). O processo foi iniciado pela adição de 100µl de solução tampão de purificação (*Direct PCR Purification Buffer*) e 1ml de resina, a 23µl das reacções de PCR. Esta mistura foi adicionada à minicoluna com o auxílio de uma seringa. Seguiu-se uma lavagem da coluna com 2ml de isopropanol a 80% e um passo de centrifugação a 16000g para remoção do isopropanol residual. O DNA foi posteriormente eluído em 50µl de água tratada com DEPC.

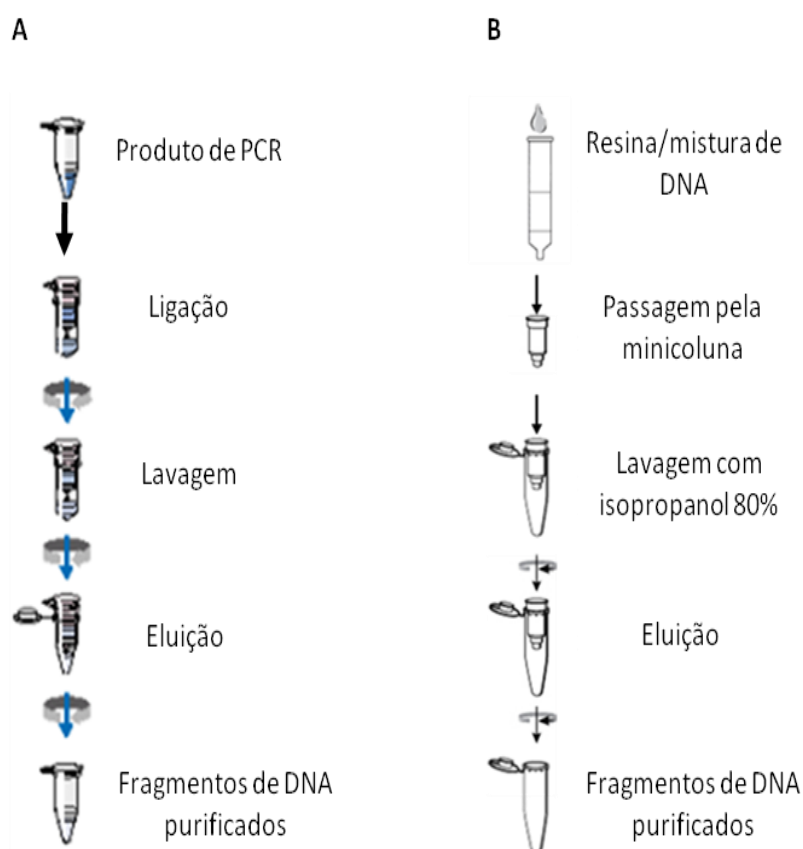


Figura 2.4. A) Esquema representativo do procedimento efectuado para purificação dos produtos de PCR com os sistemas comerciais A) *QIAquick PCR Purification Kit* e B) com *Wizard® PCR Preps DNA Purification System*. (Adaptado de *QIAquick Spin Handbook 03/2008*, Quiagen e *Technical Bulletin Wizard® PCR Preps DNA Purification System INSTRUCTIONS FOR USE OF PRODUCTS A7170, A2180, A7181 AND A7211*, Promega, respectivamente).

O rendimento dos procedimentos de purificação foi verificado através da análise, por electroforese em gel de agarose (como descrito na secção 2.6.), do DNA contido em 3µl de produto de purificação.

2.8. Clonagem molecular de fragmentos do genoma de GBV-C

Diferentes fragmentos de DNA amplificados a partir do genoma de GBV-C foram, quando necessário (antes das respectivas sequências terem sido obtidas por sequenciação de DNA), clonados em vectores plasmídicos usando *Escherichia coli* como hospedeira.

2.8.1. Preparação de células competentes de *E. coli*

O procedimento efectuado foi realizado com base no protocolo descrito por Chung e colaboradores (Chung *et al.*, 1989).

O processo foi iniciado através da obtenção uma cultura da estirpe JM109, a qual foi inoculada em 3ml de meio LB líquido (Sambrook *et al.*, 1989), a partir de uma suspensão celular (pura e concentrada) conservada a -80°C, seguida da sua incubação, durante a noite, a 37°C com agitação. Após esse período, 50µl da suspensão celular obtida foram diluídos em 15ml de meio LB (Anexo IV) tendo o crescimento celular decorrido (37°C com agitação) até ter sido atingida uma densidade óptica, medida a 600nm, a variar entre 0,3-0,5. As células bacterianas foram, então, recolhidas por centrifugação a 3300g durante 10min a 4°C, tendo-se desprezado o sobrenadante, e ressuspendido o sedimento celular em 500µl de TSS (*Transformation and Storage Solution*) frio (Anexo IV).

2.8.2. Clonagem de DNA nos vectores pGEM[®]-T e pGEM[®]-T Easy

Fragmentos de PCR purificados, correspondendo aos produtos de amplificação das regiões 5'UTR, E1/E2 e NS5A/NS5B do genoma de GBV-C, foram clonados em dois vectores plasmídicos: pGEM[®]-T (Promega, EUA) e pGEM[®]-T Easy (Promega, EUA). O vector pGEM[®]-T é obtido comercialmente sob a forma de moléculas lineares que apresentam resíduos de nucleótidos de timidina no extremo 3' em ambas as extremidades. Tal facto, não só evita a recircularização do vector, como facilita a ligação da maioria dos produtos de PCR ao mesmo. Como principais características, o vector contém os promotores de transcrição dos fagos T7 e SP6 a flanquear o sítio de clonagem múltipla, localizado na sequência de parte do gene *lacZ* que codifica o péptido α da enzima β -galactosidase. Assim, a inserção de um fragmento de DNA nesta região impossibilita a expressão de uma β -galactosidase activa, mesmo em estirpes bacterianas que codifiquem o fragmento ω (fragmento do gene *lacZ* com localização

cromossômica) desta enzima, e impedindo assim o fenómeno de α -complementação. Esta abordagem facilita o despiste de clones contendo moléculas de DNA recombinantes, os quais surgem como colónias brancas em meio contendo IPTG (isopropil- β -D-tiogalactopiranosídeo) e X-Gal (5-bromo-4cloro-3-indolil- β -D-galactopiranosídeo). O gene *bla*, que codifica a enzima β -lactamase a qual, por sua vez, confere resistência à ampicilina, também se encontra localizado neste plasmídeo. O mapa e características do vector pGEM[®]-T encontram-se esquematizados na Figura 2.5.

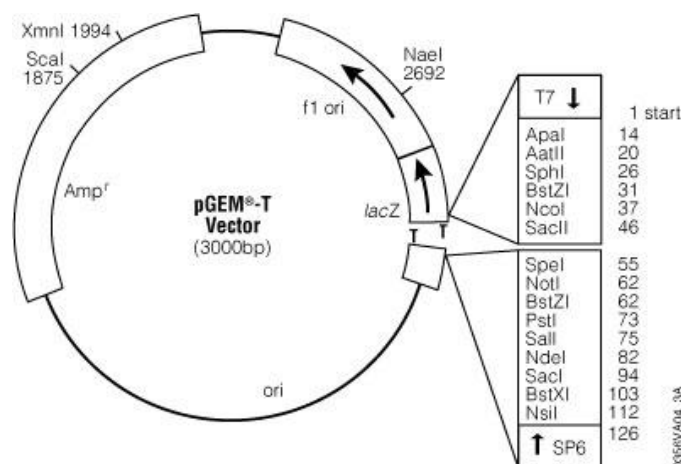


Figura 2.5. Esquema representativo do vector pGEM[®]-T (<http://www.promega.com/>).

Para ligação de DNA a este vector, preparou-se uma mistura de ligação contendo 5 μ l de solução tampão de ligação concentrada 2X (Anexo IV), 1 μ l de vector pGEM[®]-T (50ng), 3 μ l do produto de PCR purificado e 1 μ l de DNA ligase do fago T4. Estas misturas foram incubadas a 4°C durante a noite, tendo sido posteriormente utilizadas na transformação das células competentes de *E. coli*. Como alternativa foi usado outro vector de clonagem, o vector pGEM[®]-T Easy (Promega, Reino Unido/EUA). As características referentes a este vector são, em tudo, idênticas às do vector pGEM[®]-T, à excepção do facto de neste último, o local de clonagem múltipla se encontrar flanqueado de sequências reconhecidas pela endonuclease de restrição EcoRI. Desta forma, a preparação das misturas de ligação foi efectuada de acordo com o procedimento descrito para pGEM[®]-T.

2.8.3. Transformação de *E. coli* por choque térmico

Adicionou-se 5 μ l de cada mistura de ligação a 200 μ l das células ressuspendidas em meio TSS (como descrito em 2.8.1.) em microtubos frios. As suspensões assim obtidas, foram incubadas no

gelo durante 60min. Seguidamente, estas foram submetidas a um choque térmico a 42°C durante 90seg, e imediatamente recolocadas em gelo. Após 2-5 minutos no frio, foi adicionado a cada tubo um volume de 900µl de meio LB, seguido de um período de incubação de 1h a 37°C, com agitação suave. Seguidamente, distribuíram-se 200µl da suspensão bacteriana à superfície de caixas de meio LB sólido suplementado com ampicilina (100µg/ml), X-gal (80µg/ml) e IPTG (20µg/ml). As caixas foram incubadas numa estufa a 37°C, durante a noite (entre 16 a 18h) de forma a permitir o crescimento de colónias bacterianas.

2.8.4. Extracção do DNA plasmídico por lise alcalina

A extracção de DNA plasmídico foi executada através da lise alcalina de suspensões bacterianas. As colónias identificadas como possíveis clones recombinantes, obtidos pela inserção de DNA alvo nos vectores utilizados, foram repicadas, com um palito estéril, para tubos com 2ml contendo meio LB líquido suplementado com ampicilina (100µg/ml). Estes foram mantidos durante a noite a 37°C, com agitação, até à obtenção de uma cultura bacteriana saturada. Posteriormente, centrifugou-se 1,5ml da mesma a 16000g durante 2min, tendo o sedimento celular sido recuperado, após decantação do sobrenadante. As células foram, posteriormente, ressuspensas em 200µl de TEG (25mM Tris-HCl pH 8,0, 1mM EDTA, 1% glucose) frio, e as suspensões assim obtidas adicionadas de 250µl de solução de lise (0,2M NaOH, 1,5% SDS). As suspensões bacterianas foram agitadas suavemente, por inversão, para estimular o efeito de lise, e em seguida adicionadas de 250µl de acetato de potássio 3M (pH 5,4). De imediato, procedeu-se, novamente, à sua agitação por inversões sucessivas. Os tubos contendo o lisado celular foram, então, centrifugados a 16000g durante 15min, à temperatura ambiente. O sobrenadante foi transferido para novos microtubos contendo igual volume de isopropanol (aproximadamente 700µl), e o seu conteúdo homogeneizado por inversão suave. Os ácidos nucleicos precipitados foram, em seguida, colhidos por centrifugação nas mesmas condições descritas anteriormente. O sedimento de ácidos nucleicos foi lavado com 250µl de etanol e em seguida seco sob vácuo (*Concentrator 5301 - Eppendorf*, EUA). Por fim, o sedimento seco foi ressuspendido em 30µl de TE (1M Tris-HCl pH 8,0, 0,5M EDTA) com 100µg/ml RNase e incubado a 37°C, em banho-maria, durante 30minutos (para permitir uma eficiente degradação do RNA presente). Para identificação das moléculas de DNA plasmídico potencialmente recombinantes, analisou-se a migração, em gel de agarose a 1% (como descrito em 2.6.), das moléculas contidas em 2µl dos diferentes extractos de DNA plasmídico. Foi observada a diferença no padrão de migração destas

moléculas comparativamente à de um plasmídeo de referência com aproximadamente 3kb (moléculas circulares, covalentemente fechadas e super-enroladas), tendo sido seleccionadas as moléculas de plasmídeos com migração retardada para posterior análise por digestão enzimática.

2.8.5. Digestão enzimática de DNA

A identidade dos clones de DNA recombinante pretendidos foi confirmada pela sua clivagem enzimática com endonucleases de restrição. Para o efeito, foram preparadas misturas de reacção com 2µl dos extractos contendo DNA plasmídico, 2-4 unidades de enzima, 2µl de solução tampão adequada (concentrada 10X), e H₂O para um volume total de digestão de 20µl. As reacções de clivagem decorreram num banho-maria a 37°C por um período de 2h (aproximadamente). Uma vez terminada, os produtos resultantes foram analisados por electroforese em gel de agarose a 2% (Anexo IV). Os clones bacterianos contendo plasmídeos recombinantes foram re-inoculados em 2ml de meio LB com ampicilina (100µg/ml), e incubados a 37°C durante a noite, com agitação. A 1ml de cada cultura bacteriana adicionou-se 500µl de glicerol a 50% para posterior conservação a -80°C.

2.8.6. Análise de moléculas recombinantes de DNA plasmídico através da amplificação, por PCR, dos fragmentos clonados

Quando conveniente, a confirmação da natureza recombinante das moléculas de DNA plasmídico foi efectuada através da amplificação, por PCR, dos insertos de DNA nelas contidos. Foram preparadas misturas de reacção, para um volume final de 25µl, com uma mistura de reacção pronta a usar (*2X PCR Master MIX* - Fermentas, Lituânia), como descrito na secção 2.5.2. Em cada reacção foram adicionados 12,5µl de *2X PCR Master MIX*, 7,5µl de água tratada com DEPC, 2µl do *primer* directo (*primer* “universal” com sequência complementar à sequência do promotor T7), o mesmo volume de *primer* reverso (*primer* “universal” com sequência complementar à sequência do promotor do fago SP6), e 1µl da solução contendo o DNA plasmídico extraído.

As condições de amplificação utilizadas encontram-se descritas na Tabela 2.6. A reacção de amplificação foi executada no termociclador *myCycler* (Bio-Rad, EUA).

Os produtos amplificados foram, de seguida, observados após a sua separação por electroforese em gel de agarose a 1,5% (tal como descrito em 2.6.).

Tabela 2.6. Condições de reacção utilizadas nas reacções de PCR destinadas à amplificação dos fragmentos de DNA clonados.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C)	TEMPO DE DURAÇÃO
Desnaturação inicial	1	95°C	2´
Desnaturação	35	95°C	30´´
Hibridação		51°C	45´´
Extensão		72°C	1,30´
Extensão final	1	72°C	5´
Manutenção	1	4°C	∞

2.8.7. Extração de DNA plasmídico em média escala

O DNA plasmídico contido nos clones recombinantes de interesse foi extraído através do sistema comercial *Plasmid purification Midi Kit* (QIAGEN, Alemanha), seguindo as instruções do fabricante. Este protocolo foi otimizado para preparação de até 20µg de moléculas de DNA plasmídico de elevado número de cópias por célula. O DNA extraído foi eluído num volume final de 150µl de 10mM Tris-HCl, pH 8,0, detectado por electroforese em gel de agarose a 1%, e quantificado por análise espectrofotométrica (Smartspec 3000 - BioRad, EUA) (Sambrook *et al.*, 1989).

2.9. Ensaaios de PCR em Tempo Real (*Real Time PCR*)

As reacções de amplificação por PCR em Tempo Real realizadas no decurso deste trabalho foram efectuadas em microtubos de 0,2ml de parede fina (Axygen, EUA) num *Rotor-gene 6 version 6.1* (Corbett Life Science/Qiagen, EUA), tendo sido utilizados dois métodos diferentes de detecção do produto amplificado.

Um dos métodos consistiu na utilização de um fluocromo, designado por *SYBR Green I*, com capacidade para se intercalar na cadeia dupla de DNA de forma inespecífica (Figura 2.6.), emitindo, em consequência, fluorescência verde quando excitado com radiação de comprimento de onda adequado (470-500nm).

O princípio do método baseia-se, então, na detecção de fluorescência no tubo de reacção à medida que o corante se liga à cadeia dupla das moléculas de DNA amplificadas durante a reacção de PCR (Kubista *et al.*, 2006; Arikawa *et al.*, 2008).

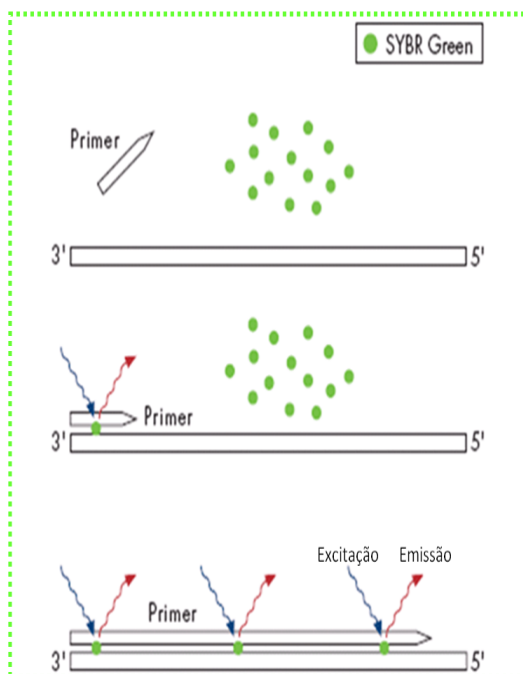


Figura 2.6. Representação esquemática de um método de PCR em Tempo Real em que a química de detecção do(s) produto(s) amplificado(s) se baseia na utilização de *SYBR Green I* (<http://www.meduni-graz.at/>).

Por sua vez, o segundo método de monitorização da acumulação dos produtos amplificados, baseou-se na aplicação de sondas oligonucleotídicas marcadas com corantes fluorescentes, conhecidas vulgarmente por sondas de hidrólise ou *TaqMan*. Estas sondas são específicas para o segmento de DNA alvo que se pretende amplificar e apresentam um fluoróforo de referência covalentemente ligado ao extremo 5', o qual é capaz de absorver radiação electromagnética num comprimento de onda específico (radiação de excitação emitida pelo equipamento) e dissipá-la na forma de luz e calor, em comprimento de onda diferente do original (Kubista *et al.*, 2006; Arikawa *et al.*, 2008).

Neste tipo de sondas, toda a radiação emitida pelo fluoróforo do extremo 5' é absorvida por uma outra molécula que a bloqueia (*quencher*), e a qual se encontra ligada de forma covalente ao extremo 3' da sonda. Desta forma, o sistema óptico do equipamento é incapaz de detectar fluorescência no tubo de reacção (acima de um valor mínimo de “fundo” - *background*). Por outro lado, se a reacção for capaz de gerar produtos de amplificação (amplicões), a sonda poderá então hibridar com a sequência alvo de que é específica, ficando exposta à actividade de

exonuclease (5'→3') da polimerase de DNA utilizada. Como consequência, a sonda é degradada a partir do seu extremo 5' resultando na separação física do fluoróforo e do *quencher*, que agora não mais será capaz de absorver a radiação emitida pelo primeiro (Figura 2.7.).

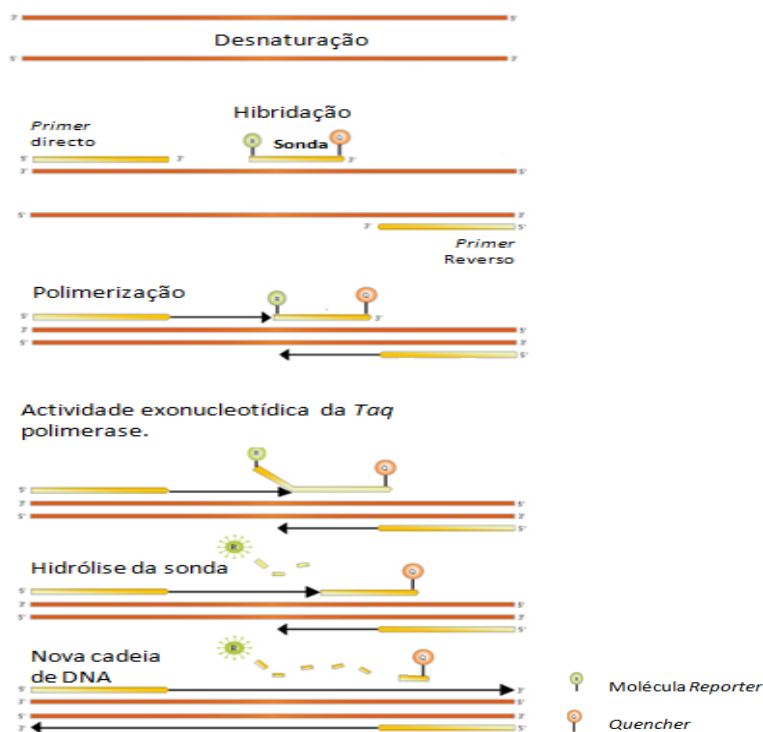


Figura 2.7. Representação esquemática da sonda *TaqMan* (adaptado de <http://www.servicexs.com>).

2.9.1. Construção de uma recta-padrão de amplificação

Um fragmento de DNA com aproximadamente 250pb, produto da amplificação por *nested*-PCR de parte da região 5'UTR (tal como descrito em 2.5.1.) a partir de uma das estirpes de GBV-C detectadas no decurso deste trabalho, foi clonado no vector pGEM[®]-T, tal como descrito em 2.8.2. O plasmídeo recombinante assim obtido, denominado por pCB1, foi extraído em média escala (ver secção 2.8.7.) e quantificado espectrofotometricamente, tendo o valor obtido, expresso em ng de DNA por unidade de volume (µl) sido convertido em número de moléculas de DNA de cadeia-dupla por unidade de volume através da fórmula, $[(\text{quantidade de DNA}(\text{ng}) * 6.022 \times 10^{23}) / (\text{tamanho do fragmento}(\text{pb}) * 1 \times 10^9 * 650)]$, disponível em <http://www.uri.edu/research/gsc/resources/cndna.html>. A construção da recta padrão para quantificação de moléculas de DNA amplificadas por PCR em Tempo Real, baseou-se na amplificação por PCR utilizando como matriz as moléculas do plasmídeo acima citado, presentes

em diluições seriadas variando entre 10^{-1} e 10^{-10} , correspondentes a um número de moléculas a variar entre $1,4 \times 10^{11}$ e 140. O procedimento experimental utilizado encontra-se descrito em 2.9.2.1 e 2.9.2.2.

2.9.2. Ensaios de PCR em Tempo Real com amostras biológicas

Para avaliação semi-quantitativa do número de equivalentes de genoma em amostras biológicas estudadas, foram realizados ensaios de PCR em Tempo Real com 9 amostras seleccionadas (com designação de A-I) de entre as utilizadas neste estudo (Anexo III). Destas, 8 são positivas para GBV-C (A-H), enquanto que uma serviu como um controlo negativo de amplificação, apesar da presença de ácidos nucleicos na amostra (amostra I). O procedimento experimental realizado encontra-se descrito de seguida (secções 2.9.2.1 e 2.9.2.2.). Nos ensaios realizados com *SYBR Green I*, as amostras estudadas foram diluídas com um factor de diluição 1:50, ao passo que aquando da utilização de sonda específica a diluição utilizada foi 1:10. Os sinais obtidos foram posteriormente analisados comparativamente aos correspondentes à curva padrão auferida.

2.9.2.1. Ensaios de PCR em Tempo Real com *SYBR Green I*

As misturas de reacção decorreram num volume final de 12,5µl, com 6,25µl de uma pré-mistura para PCR em Tempo Real contendo o corante *SYBR Green I* (*Maxima™ SYBR Green qPCR Master Mix 2X* - Fermentas, Lituânia), *Taq* DNA polimerase quimicamente modificada (inactiva à temperatura ambiente), tampão de reacção com KCl e $(\text{NH}_4)_2\text{SO}_4$ e um controlo interno (*ROX*).

Tabela 2.7. Condições de reacção utilizadas na reacção de PCR em Tempo Real com o corante *SYBR Green I*, destinada à amplificação do DNA plasmídico correspondente à região 5'UTR.

ETAPA	NÚMERO DE CICLOS	TEMPERATURA (°C)	TEMPO DE DURAÇÃO
Desnaturação inicial	1	95°C	2'
Desnaturação	45	95°C	45''
Hibridação		60°C	45''
Extensão		72°C	45''
Extensão final	1	72°C	5'
Manutenção	-	4°C	∞

A esta pré-mistura foram adicionados 2,25µl de água tratada com DEPC, 1µl de cada um dos *primers* GBV-CFor e GBVC-Rev (10pmol/µl) (Tabela 2.8.), e 2µl de DNA matriz. Para o controlo negativo, o volume de DNA foi substituído por 2µl água tratada com DEPC. As condições de reacção encontram-se representadas na Tabela 2.7.

2.9.2.2. Ensaios de PCR em Tempo Real com sonda *TaqMan*

A sonda *TaqMan*, bem como o par de *primers* utilizados nestes ensaios, foram descritos num estudo anterior (Hattori *et al.*, 2003) (Tabela 2.8.). A sonda em questão foi marcada com o fluorocromo Cy3 na extremidade 5', o qual absorve radiação no comprimento de onda $\lambda = 548\text{nm}$ e emite fluorescência a $\lambda = 562\text{nm}$. A extremidade 3' por seu lado, foi marcada com BHQ-2.

As misturas de reacção foram preparadas para um volume final de 12,5µl contendo 6,25µl de uma pré-mistura para PCR em Tempo-Real (*Dynamo probe system (qPCR) 2X* - Finnzymes, Finlândia). A este volume foram adicionados 1µl de água tratada com DEPC, 1µl de sonda (50mM final), 1µl de cada um dos *primers* (200mM final), 0,25µl de ROX (concentrado 50X) e 2µl de DNA. No controlo negativo, o volume de DNA foi substituído por 2µl água tratada com DEPC. As condições de reacção foram análogas às descritas na Tabela 2.7.

Tabela 2.8. Informação relativa ao conjunto de *primers* e sonda usados para amplificação de fragmentos de DNA de GBV-C por PCR em Tempo Real.

PRIMER (DESIGNAÇÃO)	SEQUÊNCIA (5' → 3')	TAMANHO (NT)	TM (°C) ¹	COORDENADAS (NC_001710)
GBV-Cprobe	CCCGACGTCAGGCTCGTCGTTAA ^A	24	61°C	267 – 290
GBV-CFor	ACGGTCCACAGGTGTTGG	18	53°C	228 – 245
GBV-CRev	GGACATTGAAGGGCGACG	18	53°C	337 – 354

¹A temperatura de desnaturação (*melting*) dos *primers*, foi calculada com recurso à ferramenta *Oligo Calc: Oligonucleotide Properties Calculator*, disponível em <http://www.basic.northwestern.edu/biotools/oligocalc.html>. NT indica nucleótidos, enquanto que TM indica temperatura de desnaturação (do inglês *temperature of melting*).

2.9.2.3. Ensaios de *heminested*-PCR com amostras testadas por PCR em Tempo Real

Nas situações que assim o justificaram, procedeu-se a uma reacção de *heminested*-PCR para amplificação de um fragmento do genoma viral que inclui parte da 5'UTR ($\approx 250\text{nt}$), a partir de

algumas amostras biológicas testadas nos ensaios de PCR em Tempo Real. Para tal, recorreu-se ao sistema *puReTaq Ready-To-Go PCR Beads* para um volume final de 25µl. Resumidamente, na primeira reacção usaram-se 5µl de cDNA matriz, com 2µl de cada um dos *primers* (800mM final) UTRForO e E1rev2, adicionados a 16µl de H₂O livre de nucleases juntamente com uma esfera liofilizada. Finda a primeira reacção de amplificação, 2µl do produto amplificado foram adicionados a um novo tubo de reacção, com 19µl de H₂O livre de nucleases 2µl de cada um dos *primers* (800mM final) UTRForO e E1rev1, e submetidos à segunda etapa de PCR. Após observação dos amplicões em gel de agarose a 1,5%, as respectivas amostras foram posteriormente sequenciadas.

2.10. Sequenciação de DNA

Moléculas de DNA purificadas foram sequenciadas pela empresa STAB VIDA (Portugal), por recurso à técnica de terminação da polimerização em cadeia com didesoxi-ribonucleótidos, descrita inicialmente por Sanger (Sanger *et al.*, 1977). Foi efectuada a sequenciação, em ambos os sentidos, dos fragmentos de DNA purificados, tendo sido, para tal, usados como iniciadores, os *primers* utilizados nas segundas reacções de PCR. As moléculas de DNA plasmídico recombinante, contendo fragmentos de PCR clonados, foram sequenciadas, em ambas as cadeias, por recurso aos *primers* universais T7 e SP6, complementares às sequências que flanqueiam o local de clonagem múltipla dos vectores plasmídicos utilizados.

O produto obtido nas reacções de *heminested*-PCR descritas em 2.9.2.3., foi sequenciado com o *primer* UTRForO.

2.11. Análise bioinformática

2.11.1. Análise das sequências nucleotídicas/proteicas

Como indicado em 2.10., os produtos de PCR obtidos foram sequenciados, em ambos os sentidos, com os *primers* utilizados na sua amplificação. Em consequência, foram obtidas 2 sequências para cada uma das moléculas analisadas. Desta forma, para cada uma das amostras/região do genoma do GBV-C em estudo, foi construída uma sequência de consenso por recurso à aplicação CAP (*Contig Assembly Program*) do programa *BioEdit Sequence Alignment Editor* 7.0.9.0 (Hall, 1999). Cada uma destas sequências foi corrigida manualmente e

inicialmente analisada com o programa BLAST (disponível em <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para confirmação da sua identidade.

Para análise das sequências putativas de resíduos de aminoácidos, as sequências nucleotídicas de GBV-C foram traduzidas com o programa bioinformático *GeneRunner* versão 3.05 (aplicação informática de utilização livre obtida a partir de <http://www.generunner.net/>). A construção de alinhamentos múltiplos de sequências de resíduos de aminoácidos foi efectuada com o programa MAFFT *version 6* (referido em 2.4.). O valor da entropia de Shannon, considerado como medida de variabilidade existente nos alinhamentos múltiplos construídos, foi calculado através da utilização da ferramenta ENTROPY-TWO (acessível em <http://www.hiv.lanl.gov/content/sequence/ENTROPY/>).

As distâncias genéticas intra e intergenótipos, foram determinadas com o programa MEGA 4.0., e os cálculos das taxas de substituições sinónimas (d_S) e não sinónimas (d_N), efectuados de acordo com o método descrito por Nei e Gojobori (Nei e Gojobori, 1986) através da utilização do programa informático SNAP (*Synonymous Non-synonymous Analysis Program*) disponível em <http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>.

Por outro lado, a ferramenta PCOORD (disponível em <http://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html>), através da construção de alinhamentos múltiplos com as sequências em análise, foi utilizada para identificação/distinção de putativos sub-grupos filogenéticos formados.

2.11.2. Análise filogenética com base em alinhamentos de sequências nucleotídicas

Os alinhamentos múltiplos de sequências nucleotídicas obtidas no decurso deste trabalho foram construídos com o programa MAFFT *version 6* e a sua edição feita manualmente no programa *BioEdit Sequence Alignment Editor 7.0.9.0* (ver secção anterior).

As árvores filogenéticas foram elaboradas de acordo com três métodos de reconstrução filogenética comumente utilizados: método de junção-de-vizinhos (*Neighbor-Joining - NJ*), análise *Bayesiana* com base no processo aleatório de cadeias de Monte-Carlo (*Markov Chain Monte Carlo – MCMC*) e no critério de Máxima Verosimilhança. As matrizes de distâncias genéticas, que serviram de base à construção de árvores filogenéticas pelo método de NJ, foram calculadas com base no algoritmo de Kimura a dois parâmetros (K-2P) (Kimura 1980). A robustez das topologias sugeridas foi avaliada por análise de *bootstrap* de 1000 replicados dos

dados originais. Todos estes cálculos foram executados mediante a utilização do programa MEGA 4.0 (Tamura *et al.*, 2007). Alternativamente, as análises de inferência filogenética baseada em análise Bayesiana foram executadas ao longo de 6×10^6 gerações (usando o programa *MrBayes* v3.0b4, Huelsenbeck e Ronquist, 2001) tendo por base o modelo GTR (*General Time Reversal*). Por outro lado, as taxas de substituição nucleotídica, frequências nucleotídicas, a taxa de heterogeneidade estimada usando a distribuição γ para as posições variáveis e a frequência de posições invariáveis, foram estimadas com o programa *MrModeltest*2.3. (Nylander 2004).

As árvores filogenéticas obtidas com base no critério de máxima verossimilhança foram construídas através da utilização do programa *PAUP* 4.0b10 (Swofford, 2003), usando as taxas de substituição nucleotídica e frequências de nucleótidos de um modelo GTR, assim como os parâmetros α (da distribuição γ) e a frequência de posições invariáveis estimadas a partir das sequências alinhadas usando o programa *Modeltest*3.7 (Posada e Crandall 1998).

Foram considerados significativos os agrupamentos observados nas árvores filogenéticas de *NJ* que apresentam valores de *bootstrap* iguais ou superiores a 75% (Felsenstein, 1985; Higgs e Attwood, 2005), ou valores de probabilidade *a posteriori* $>0,9$ (análise Bayesiana).

NOTA:

Os modelos de reconstrução filogenética escolhidos utilizam algoritmos matemáticos de inferência estatística divergentes entre si, com aumento do grau de complexidade (MCMC>ML>NJ). O método de junção-de-vizinhos (*Neighbor-Joining* - NJ) consiste num algoritmo de adição que procura, sequencialmente, encontrar “vizinhos” de forma a que estes minimizem o comprimento total dos seus ramos de uma única árvore filogenética (Saitou e Nei, 1987; Hall, 2005; Gascuel e Steel, 2006). Este processo filogenético teve por base a obtenção de distâncias genéticas corrigidas com base no modelo evolutivo de Kimura a dois parâmetros (Kimura 1980), que propõe dois parâmetros variáveis se substituições nucleotídicas, consoante as mutações em causa se tratem de transições ou transversões, e frequências uniformes dos 4 nucleótidos. Embora menos robusto que as demais metodologias utilizadas, apresenta como vantagem uma grande rapidez computacional e, por isso, é frequentemente o método de reconstrução filogenética mais utilizado. Neste caso, o teste de confiança filogenético que assiste a fidelidade da topologia resultante é tido pelo cálculo de valores de *bootstrap* (percentagem do número de vezes que um dado agrupamento se manteve no número de réplicas efectuadas) associados aos ramos, sendo considerados significativos os superiores a 75% (Felsenstein, 1985; Higgs e Attwood, 2005).

Por outro lado, os métodos MCMC e ML assentaram, ambos, no modelo evolutivo GTR (*General Time Reversible*), distinto do descrito anteriormente, ao considerar a existência de diferentes taxas de substituição nucleotídica bem como de frequências nucleotídicas tratando-se, como tal, de um modelo com grande número de parâmetros variáveis, permitindo uma melhor adequação a diferentes situações de evolução de sequências de DNA (Li *et al.*, 1985). O modelo utilizado incluiu, ainda, a introdução de um parâmetro correspondente à frequência de posições nucleotídicas invariáveis, assim como um parâmetro α de uma distribuição γ (*gamma*), a qual descreve a heterogeneidade na distribuição das taxas de substituição nucleotídica ao longo de uma sequência. A distribuição *gamma*, é uma função contínua de densidade de probabilidade muito flexível, cuja forma pode variar de acordo com o parâmetro α (especifica a amplitude da variação da taxa de substituição por cada local). As taxas de mutação em diferentes locais são consideradas como variáveis aleatórias obtidas a partir da função *gamma* de parâmetro α . Para

valores elevados de α o modelo *gamma* aproxima-se da independência de locais, ou seja, a maioria dos locais evoluem a uma taxa constante, enquanto que para valores baixos, existe uma enorme variabilidade das taxas de mutação, com uma grande percentagem de locais com taxas reduzidas e uma pequena percentagem de locais com taxas elevadas (Li *et al.*, 1985).

A reconstrução filogenética por Máxima Verosimilhança baseia-se no cálculo de um valor de probabilidade (verosimilhança/*likelihood*-L) de que, com base num modelo molecular assumido, as sequências analisadas tenham evoluído segundo a árvore escolhida. A pesquisa de árvores filogenéticas (no espaço possível) com base neste critério procura, então, encontrar a árvore que confere a maior probabilidade ao alinhamento de sequências observado (Felsenstein, 1981). No contexto filogenético, a verosimilhança representa então a probabilidade de se observarem os dados tendo por condições base uma árvore e um modelo de evolução de DNA. Por seu lado, na abordagem Bayesiana, que se baseia na análise de um grande conjunto de árvores com probabilidades aproximadamente iguais, todas as inferências passam a ser feitas tendo em conta as probabilidades *a posteriori* dos parâmetros evolutivos e dos *taxa* agrupados (Rannala e Yang, 1996). Neste procedimento, várias árvores filogenéticas são geradas em sucessão numa cadeia de Markov por um processo aleatório de Monte-Carlo. Uma das grandes vantagens relaciona-se com o facto de conferir probabilidade aos parâmetros estimados e aos agrupamentos de árvores filogenéticas, para além de permitir uma pesquisa mais exaustiva do espaço de árvores possíveis.

A presença de possíveis recombinantes foi pesquisada por *bootscanning* através do programa bioinformático *Simplot* (Lole *et al.*, 1999), recorrendo para tal, a um conjunto de sequências de referência alusivas aos 6 genótipos reconhecidos de GBV-C, disponíveis nas bases de dados (EMBL/GenBank/DDBJ). Esta análise foi efectuada considerando árvores filogenéticas de NJ com distâncias genéticas baseadas no modelo de Kimura a dois parâmetros, estimadas numa janela inicial de análise da sequência de 180pb, seguida de incrementos de 10pb e 1000 replicados.

2.11.3. Análise estatística

Uma análise da associação entre variáveis seleccionadas foi realizada com o teste estatístico de χ^2 (com correcção de *Yates*) e os resultados obtidos aplicados na elaboração de tabelas de contingência. Os valores com $P < 0,05$ foram considerados estatisticamente significativos e submetidos ao teste exacto de *Fisher* bilateral. Os testes referidos encontram-se disponíveis em <http://www.physics.csbsju.edu/stats/Index.html>.

A significância estatística das diferenças de entropia de Shannon observadas entre grupos de sequências analisadas, foi testada de acordo com o processo aleatório de Monte Carlo (com reposição, disponível em <http://www.hiv.lanl.gov/content/sequence/ENTROPY>), com 1000 repetições aleatórias, e onde 5, foi o número de amostragens aleatórias que se considerou poderem apresentar uma maior diferença de entropia do que o conjunto de dados analisados (ou seja com $P \leq 0,005$).

2.11.4. Previsão de formação de estruturas secundárias em moléculas de RNA

A previsão do tipo e estabilidade (inferida a partir dos valores da energia livre de *Gibbs*) das estruturas secundárias formadas em moléculas de RNA compreendendo a região 5'UTR/E1 do genoma de GBV-C foram efectuadas com base na análise das respectivas sequências nucleotídicas e recurso ao algoritmo *RNAfold* (disponível em <http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>).

3. RESULTADOS



3. RESULTADOS

3.1. Detecção de genomas do vírus GBV-C em plasma humano (virémia)

A detecção de virémia de GBV-C foi realizada tendo por base a amplificação, *in vitro*, de sequências genómicas virais. Esta pressupõe uma reacção de transcrição reversa, uma vez que o genoma viral é constituído por uma molécula de RNA, seguida por uma reacção de amplificação de parte desta, em dois passos, por *nested*-PCR. A secção do genoma escolhida como “alvo” para a detecção pretendida, correspondeu à região 5'UTR, a mais conservada do genoma deste vírus.

A análise dos produtos obtidos como resultado do primeiro passo do processo de amplificação, não revelou a presença observável de DNA em géis de agarose (Figura 3.1). Em consequência, verificou-se ser necessária a utilização de uma segunda reacção de amplificação, justificando-se, assim, a abordagem experimental escolhida.

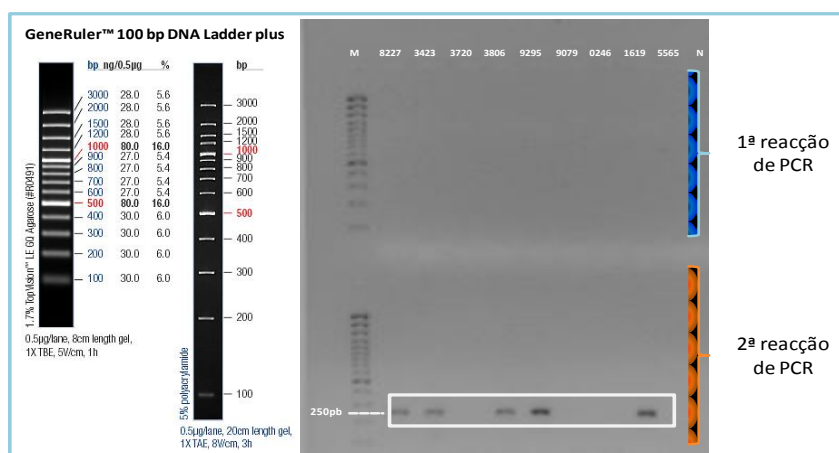


Figura 3.1. Observação dos produtos de PCR resultantes da amplificação da região 5'UTR após a sua separação por electroforese em gel de agarose. Identificação das pistas: 8227, 3423, 3720, 3806, 9295, 9079, 0246, 1619 e 5565 - amostras clínicas em estudo; N - controlo negativo (cDNA matriz substituído por água); M - marcador de massa molecular, representado à esquerda (*GeneRuler™100bp DNA Ladder plus*). O fragmento, de aproximadamente 250pb, correspondente ao produto de amplificação pretendido, encontra-se indicado por um rectângulo branco (amostras: 8227, 3423, 3806, 9295 e 1619)

Um fragmento de DNA com tamanho de aproximadamente 250pb (coordenadas 224-476 na sequência de genoma completo NC_001710) foi, deste modo, amplificado, tal como revelado,

em algumas amostras, pela separação por electroforese em gel de agarose a 1,5% dos produtos da segunda reacção de amplificação (Figura 3.1).

Um dos produtos de amplificação (amostra 1642), escolhido aleatoriamente, depois de purificado como descrito em 2.7. (Material e Métodos), foi sequenciado de forma a permitir a confirmação da origem do DNA obtido. A sequência nucleotídica daí resultante, depois de editada manualmente, foi submetida a uma pesquisa automática de homologia com o genoma de GBV-C, por recurso à ferramenta bioinformática *BLASTn*, identificando-a, indubitavelmente, como sendo a região 5'UTR do genoma de GBV-C.

Usando a estratégia acima descrita, no total das 214 amostras analisadas, foi detectada a presença de RNA genómico de GBV-C em 87 delas (Anexo I), indicando uma taxa de virémia na população estudada de 40,65%. Destas, 54 (62,10%) foram obtidas de indivíduos do sexo masculino (média de idade=32,8) e 33 (37,93%) do sexo feminino (média de idade=36,8) (Figura 3.2 A). A maioria (72,40%) dos indivíduos onde foi detectada virémia de GBV-C tinha nacionalidade Portuguesa (Figura 3.2 B), enquanto que 78,16% reportaram serem utilizadores de drogas injectáveis (68 indivíduos).

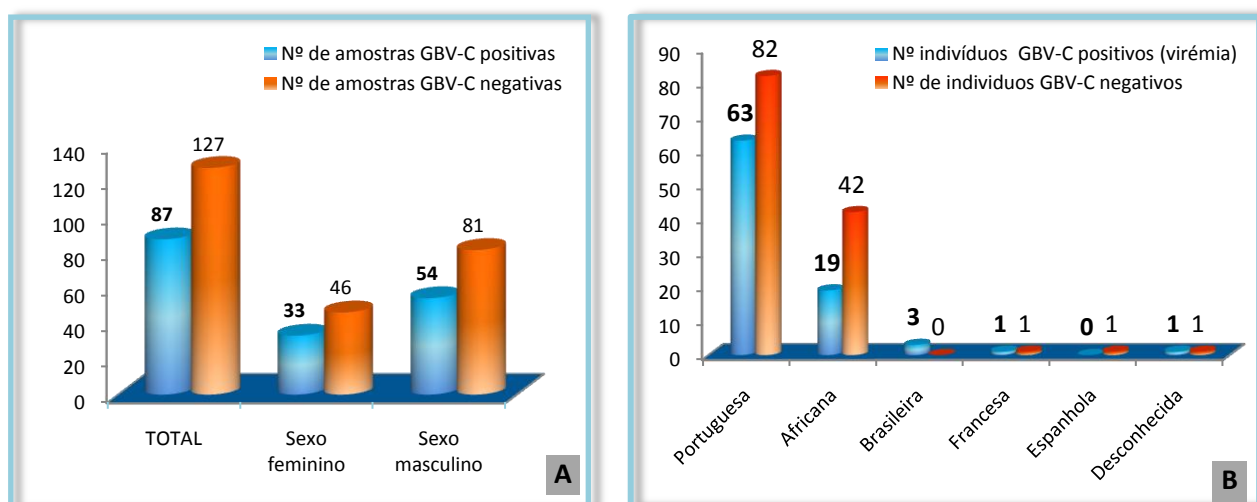


Figura 3.2. Representação gráfica do número total de indivíduos (valores absolutos), onde foi detectada a presença de RNA de GBV-C. Discriminação dos resultados obtidos por sexo **A**) e por nacionalidade **B**).

A presença de co-infecções com outros vírus de transmissão parentérica, nomeadamente o HIV e/ou o HCV) foi detectada em 17 indivíduos seropositivos para HIV (19,54%) e 14 seropositivos para HCV (16,09%). A presença de infecções triplas foi registada em 41,38% (n=36) das amostras analisadas, tendo em 22,99% das mesmas (n=20) sido detectada infecção exclusiva (monoinfecção) com GBV-C (Figura 3.3 A). Do total de indivíduos co-infectados com HIV (n=55, 63,22%), 65,50% (n=36) reportaram terem sido infectados como resultado do consumo

de drogas ilegais por via intravenosa, enquanto que 32,70% (n=18) referiram terem sido infectados pelo HIV como consequência de contactos sexuais não protegidos. Em 1,80% (n=1), e tal como indicado no capítulo anterior, a via de transmissão não foi indicada no inquérito epidemiológico efectuado. Considerando os dados disponíveis (para um total de n=43 amostras clínicas), a média da carga viral relativa ao HIV foi estimada em $3,0 \times 10^4$ (variando entre <50 e $14,5 \times 10^5$ cópias de genoma viral/ml) e a contagem de células $CD4^+$ em $4,0 \times 10^2$ (variando entre 14 e $1,1 \times 10^3$ células/ μ l).

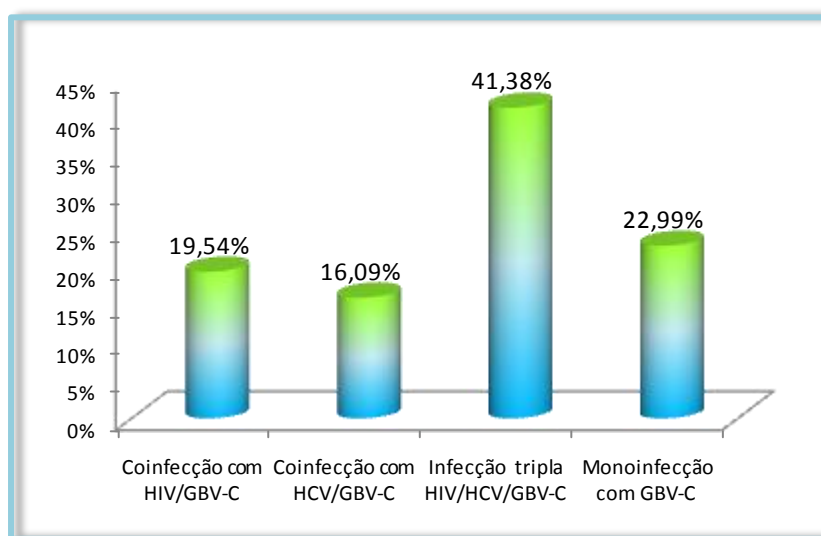


Figura 3.3. Representação gráfica do número de indivíduos, em percentagem, onde foi detectada a presença de co-infeções com os vírus HIV e HCV, bem como monoinfecção com GBV-C.

Os resultados acima referidos foram submetidos a uma análise estatística (teste de χ^2 com correcção de Yates), tendo sido pesquisada a associação da detecção de virémia de GBV-C às variáveis, sexo, idade, etnia (Africana/não Africana), toxicodependência, número de células $CD4^+$ (≤ 250 vs. >250 células/ μ l) e carga viral ($\leq 10^3$ vs. $>10^3$ RNA cópias/ml) (para a população co-infectada com HIV). Os valores obtidos demonstraram uma associação estatisticamente significativa ($P < 0,05$) entre a presença de virémia de GBV-C e a idade dos indivíduos infectados (≥ 25 anos, $P = 0,03$). A utilização de drogas injectáveis foi também associada à presença de virémia de GBV-C ($P = 0,04$). Nos restantes casos não se verificou dependência entre as variáveis em estudo.

3.2. Amplificação das regiões codificantes 5'UTR/E1, E1/E2 e NS5A/NS5B

3.2.1 Análise da região 5'UTR/E1

A análise da região 5'UTR/E1 (definida na secção 2.5.2 do capítulo anterior) envolveu 30 amostras clínicas onde foi detectada virémia relativa ao GBV-C (Anexo II), sendo que em 27 destas (90,00%), a região alvo pretendida foi amplificada sob a forma de um fragmento com aproximadamente 1,2kpb. Este incluiu parte da 5'UTR (\approx 230nt) e a região que codifica grande parte da glicoproteína E1 da superfície (invólucro) viral (\approx 940nt) (ver Figura 3.9, secção 3.2.2). O protocolo de amplificação para a segunda reacção de PCR (ver secção 2.5.2) resultou de um processo de optimização das condições de reacção, nomeadamente no que diz respeito às temperaturas de hibridação dos *primers* utilizadas (Figura 3.4 A). Nestas reacções, a temperatura de hibridação foi alterada, nos primeiros 25 ciclos de amplificação, de 55°C para 58°C, e nos últimos 10 ciclos aumentada de 55°C para 60°C (Tabela 2.5). No entanto, dado que um dos *primers* interno é degenerado (E1rev3), foram previamente testadas, como alternativa, temperaturas de hibridação intermédias variando entre os 55°C e os 65°C. As condições optimizadas revelaram uma melhoria significativa na especificidade do produto amplificado (Figura 3.4 B).

Para os casos onde se verificou a presença de vários fragmentos resultantes de amplificações inespecíficas, os amplicões (ou produtos de amplificação) obtidos na primeira reacção de PCR, foram diluídos 20X e submetidos ao procedimento descrito em 2.5.2. No entanto, o rendimento de amplificação para algumas das amostras analisadas foi baixo, tal como revelado pela presença, em gel de agarose, de bandas correspondentes a produtos de amplificação com fraca intensidade (Figura 3.4 A). Este contratempo foi ultrapassado pela utilização, na segunda reacção de PCR, de um sistema de amplificação mais robusto (método *puReTaq Ready-To-Go PCR Beads*) do que o inicialmente utilizado (mistura de reacção pronta a usar: *2X PCR Master MIX* – Fermentas, Lituânia) e sem diluição do DNA matriz.

Nas demais amostras (3 no total, 10,00%) não foi obtido o produto de amplificação desejado. Para estes três casos foram alteradas as condições de amplificação, por aplicação de um gradiente de temperaturas de hibridação na segunda reacção de PCR, variando entre os 55 e

65°C, com aumento da quantidade do DNA matriz, de 1µl para 2-4µl (volume da primeira reacção de PCR). No entanto, ou não foram conseguidos amplicões ou estes foram obtidos, sistematicamente, com um baixo rendimento (tal como indicado para a amostra 7888, Figura 3.4 A).

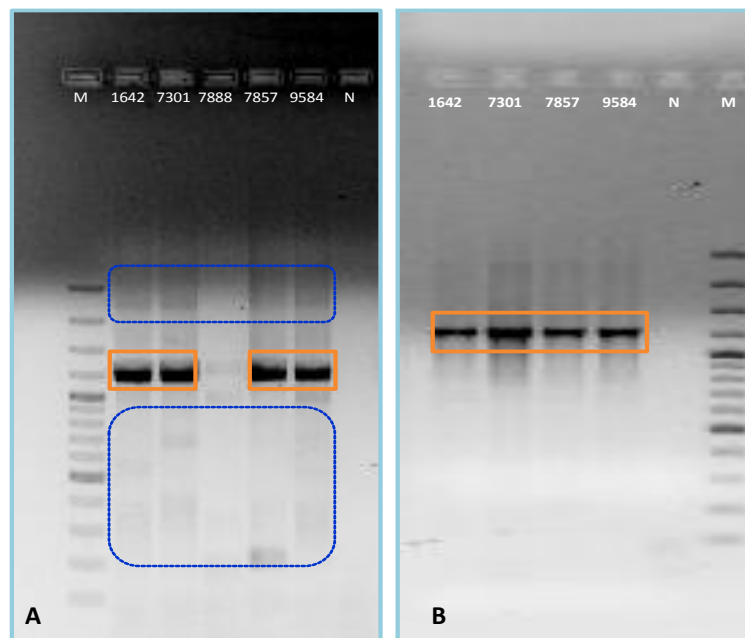


Figura 3.4. Observação em gel de agarose dos produtos relativos à amplificação por PCR da região 5'UTR/E1 do genoma de GBV-C. Os números indicados em 5 das pistas do gel correspondem às amostras de DNA em estudo; M - marcador de massa molecular (*GeneRuler™ 100bp DNA Ladder plus*); N - controlo negativo. Os fragmentos correspondentes ao produto de amplificação pretendido encontram-se indicados por um rectângulo laranja. **A)** Análise dos resultados de amplificação obtidos na segunda reacção de PCR em condições sub-óptimas, com temperatura de hibridação a 62°C durante 35 ciclos; Presença de produtos inespecíficos demarcados com pontilhado azul; **B)** Produtos de amplificação obtidos na segunda reacção de PCR com alteração das condições de reacção. Neste caso a temperatura de hibridação durante os primeiros 25 ciclos foi de 58°C, tendo nos 10 ciclos seguintes sido elevada para 60°C.

3.2.1.1 Análise da sequência da região 5'UTR/E1 do genoma de GBV-C: possível impacto da formação de estruturas secundárias

Seis dos produtos específicos resultantes das reacções de amplificação da região 5'UTR/E1 foram purificados e sequenciados. A sequenciação foi realizada em ambos os sentidos com os *primers* UTRForI e Elrev3 tendo, por consequência, sido obtidas 2 sequências para cada um dos fragmentos analisados. As sequências, que resultaram da utilização do *primer* Elrev3, apresentaram aproximadamente 760nt, e corresponderam ao resultado esperado. Pelo contrário,

as sequências originadas com o *primer* UTRForI terminaram precocemente. A mais pequena das sequências ocasionadas cobriu ≈ 140 nt do fragmento original, enquanto que a mais longa apresentava não mais de 335nt. Das seis sequências analisadas, apenas uma (correspondente à amostra 4088) foi excepção a esta situação, na medida em que ambos os resultados de sequenciação, foram de boa qualidade. Em consequência, com esta amostra foi construída uma sequência de consenso, a qual depois de editada manualmente, resultou numa sequência com 1030nt.

No sentido de melhorar os resultados auferidos por recurso à sequenciação com o *primer* UTRForI, foi desenhado um novo *primer*, designado E1rev4 (Tabela 2.2), o qual, em combinação com o primeiro, deveria permitir a obtenção de sequências que abrangessem todo o fragmento em análise. Para testar esta hipótese, 4 destes fragmentos, referentes às amostras 1642, 0594, 3273, 9709, foram novamente sequenciados (com substituição do *primer* UTRFor I pelo *primer* E1rev4). Analogamente ao sucedido anteriormente, os resultados de sequenciação revelaram sequências com apenas ≈ 350 nt, terminando, prematuramente, na mesma região do genoma de GBV-C onde terminaram as sequências obtidas com o *primer* UTRForI (Figura 3.5).

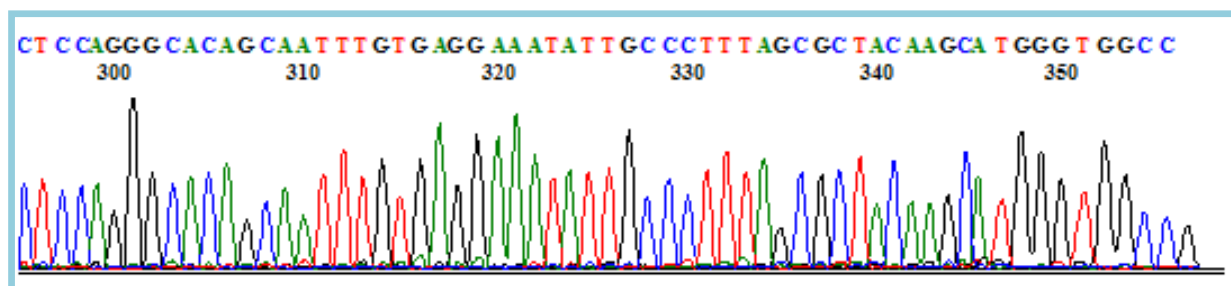


Figura 3.5. Exemplo de um dos cromatogramas que resultou da sequenciação das amostras com o *primer* E1rev4 (visualizado com o programa *BioEdit* 7.0.9.0). A Figura ilustra a sequência correspondente à amostra 0594 (por questões gráficas, e para fácil visualização, encontra-se apenas representada a parte final da sequência, a qual cessa aos 355nt).

Em consequência destes resultados, foi decidido não dar continuidade à sequenciação dos restantes 21 fragmentos 5'UTR/E1 anteriormente amplificados (secção 3.2.1).

Com base em estudos anteriores (Turner *et al.*, 2004) é sabido que na região 5'UTR do genoma de GBV-C, a existência de inúmeras sequências repetidas e invertidas promove a formação de estruturas secundárias no RNA viral. Tal facto poderia justificar a paragem da polimerase de DNA durante as reacção de sequenciação quando, durante o processo de amplificação cíclica, as

moléculas de DNA em cadeia simples tendessem a assumir estruturas secundárias e terciárias. Desta forma, foi efectuada uma análise dos resultados de sequenciação acima descritos tendo em consideração o possível impacto da ocorrência de estruturas secundárias na região em estudo (5'UTR/E1).

O processo teve início com base num alinhamento múltiplo das 5 sequências nucleotídicas obtidas com o *primer* UTRFor, juntamente com a sequência consenso 4088, as quais foram comparadas com a sequência de referência de genoma completo, NC_001710 (Figura 3.6).

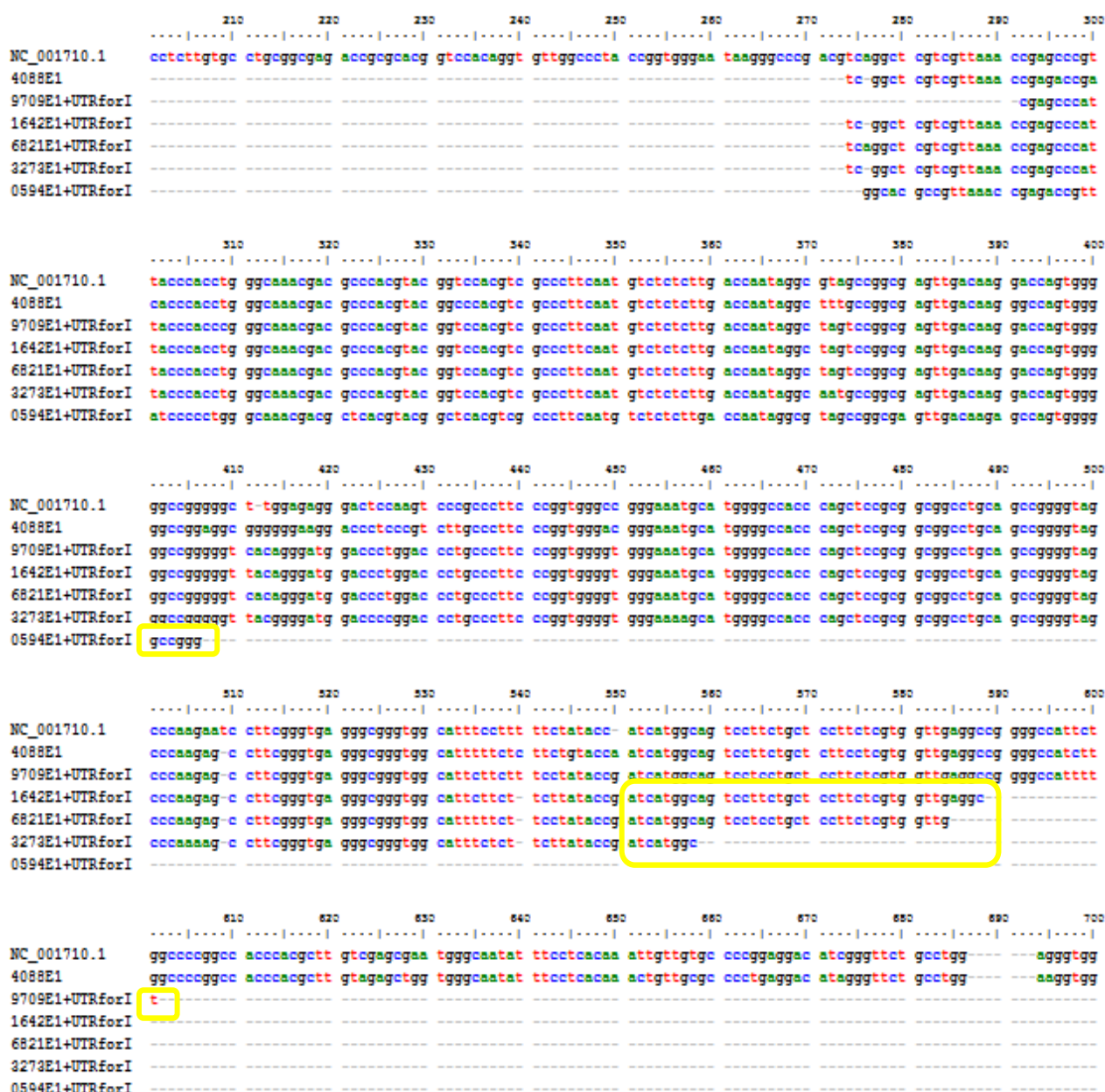


Figura 3.6. Alinhamento múltiplo das sequências referentes às amostras 9709, 1642, 6821, 3273, 0594, obtidas com o *primer* UTRForI, e da sequência consenso 4088 com a sequência de referência NC_001710. As regiões onde as reacções de sequenciação terminaram precocemente encontram-se identificadas por rectângulos amarelos.

A posição dos locais de paragem das reacções de sequenciação, relativos à sequência NC_001710, (ver Tabela 3.1 A), definidos pela posição onde, nos cromatogramas, foi registada uma perda abrupta de sinal fluorescente, “coincidem” numa região do genoma viral onde, em estudos anteriormente descritos por Simons (Simons *et al.*, 1996) (Figura 3.7 A) e por Thurner e colaboradores (Thurner *et al.*, 2004) (Figura 3.7 B), se verifica a formação de estruturas secundárias, nomeadamente ao nível das regiões designadas por HG6 e HG7 (Tabela 3.1 B).

Tabela 3.1. A) Posição nucleotídica do genoma de GBV-C (sequência NC_001710) onde terminaram precocemente as cinco amostras em análise. **B)** Reconhecimento de potenciais estruturas secundárias na região E1 e 5'UTR de GBV-C (adaptado de Thurner *et al.*, 2004).

A	Posição		B	Posição		
	Amostra	NC_001710		Designação	NC_001710	Região do genoma
	9709	601		HG1	2-28	5'-UTR
	1642	587		HG2	56-124	5'-UTR
	6821	583		HG3	277-297	5'-UTR
	3273	557		HG4	309-325	5'-UTR
	0594	408		HG5	334-398	5'-UTR
				HG6	406-433	5'-UTR
				HG7	576-614	"nucleocápside"
				HG8	646-721	E1
				HG9	828-850	E1

Com recurso à ferramenta bioinformática *RNAfold* (disponível em <http://rna.tbi.univie.ac.at/>), e tendo por base a indicação do intervalo da secção nucleotídica correspondente às referidas regiões (HG7 e HG6), foram previstas as respectivas estruturas secundárias.

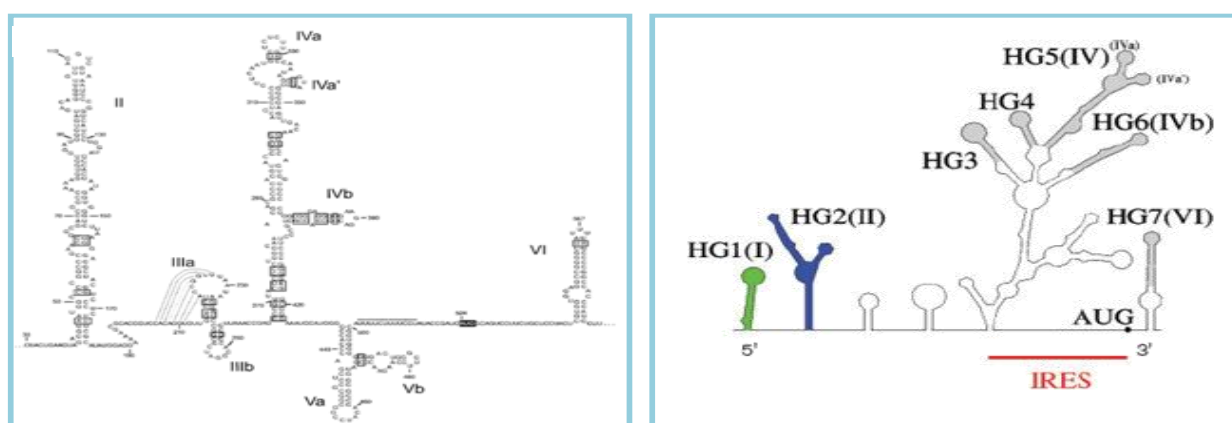


Figura 3.7. Representação esquemática das estruturas secundárias obtidas para a região 5'UTR de GBV-C em estudos anteriores. **A)** Modelo preliminar das estruturas secundárias de RNA inferidas por Simons e colaboradores. **B)** Estruturas secundárias reportadas posteriormente por Thurner e colaboradores (1-556nt). Entre parêntesis está assinalada a correspondência com as estruturas previstas pelo estudo de Simons.

A partir dos resultados obtidos (Figura 3.8), foi possível identificar a posição exacta onde terminaram as sequências da maioria das amostras em análise.

Neste sentido, prevendo a estrutura secundária correspondente à região HG6, é exequível distinguir a posição nucleotídica exacta onde terminou a sequência correspondente à amostra 0594 (Figura 3.8 A). Da mesma forma, os pontos de paragem das sequências relativas às amostras 6821, 1642 e 9709 por comparação com a previsão da estrutura secundária HG7 obtida por Thurner (Thurner *et al.*, 2004), incidiram também sobre a estrutura em gancho de cabelo (*stem-loop*) formada (Figura 3.8 B).

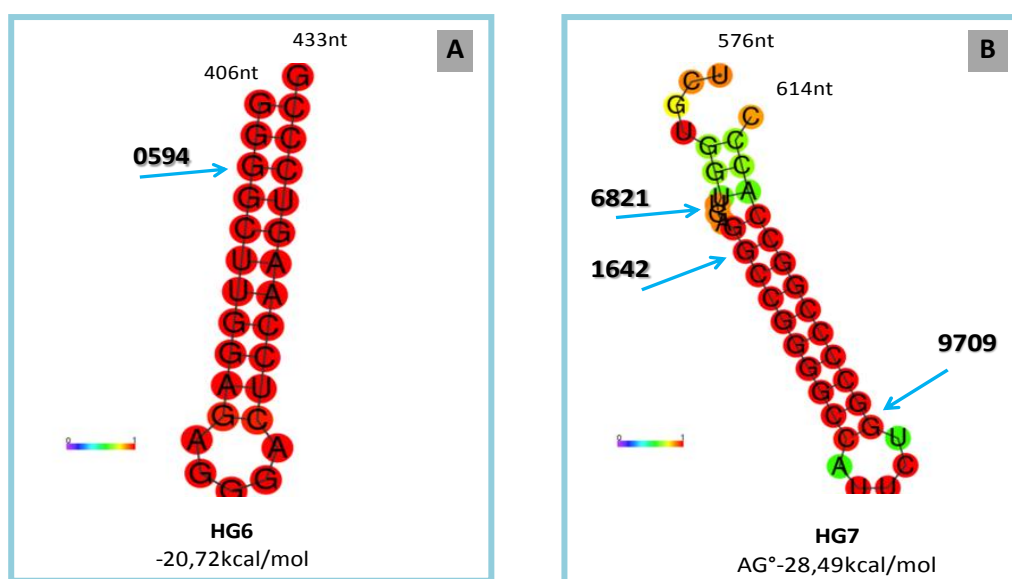


Figura 3.8. Representação gráfica das estruturas secundárias obtidas para as regiões HG6 (A) e HG7 (B) por analogia com o estudo de Thurner e colaboradores (Thurner *et al.*, 2004), com recurso à ferramenta RNAfold. As setas apontam as posições nucleotídicas correspondentes aos locais onde ocorreu paragem da reacção de sequenciação para as amostras indicadas. A barra colorida é representativa da probabilidade de emparelhamento dos pares de bases.

A avaliação da estabilidade termodinâmica das estruturas previstas, a qual está directamente relacionada com a probabilidade de formação da mesma, baseou-se no cálculo dos respectivos valores de energia livre de Gibbs (ΔG^0), (Figura 3.8). Assim ambas as regiões analisadas, tal como anteriormente evidenciado por Thurner e Simons, adquiriram conformações altamente estáveis suportadas por valores de ΔG^0 negativos, bem como uma probabilidade associada ao emparelhamento de bases maioritariamente de 1.00 (Figura 3.8).

No entanto, relativamente à amostra 3273, não existe uma consonância exacta com a formação de estruturas secundárias descritas na literatura que incluam o local de paragem da polimerase de

DNA na reacção de sequenciação referente a esta amostra. Contudo, pode inferir-se que a posição 557nt (local de terminação da sequência alvo) se situa entre as regiões assinaladas como HG6 e HG7 onde se denota, por observação da Figura 3.7, a formação de potenciais ansas (*loops*) as quais, nos fragmentos UTR5'/E1 referentes a esta amostra, poderão ter sido suficientemente estáveis para impedir a progressão da polimerase de DNA durante a reacção de sequenciação.

3.2.2 Análise das regiões E1/E2 e NS5A/NS5B

As porções da sequência codificante do genoma de GBV-C, definidas em 2.5.3 como E1/E2 e NS5A/NS5B, foram amplificadas, por *nested-PCR*, a partir de 45 (51,72%) amostras seleccionadas aleatoriamente (Anexo II) onde, previamente, fora detectada virémia de GBV-C. O fragmento que inclui parte da região que codifica as glicoproteínas E1 (≈ 240 nt) e E2 (≈ 920 nt, Figura 3.9) da superfície viral foi amplificado na totalidade das amostras seleccionadas. Por seu lado, o fragmento correspondente à região NS5A/NS5B (cobrindo ≈ 125 nt da região NS5A e ≈ 1060 nt da região NS5B, Figura 3.9), foi obtido a partir de 41 amostras seleccionadas (91,11%).

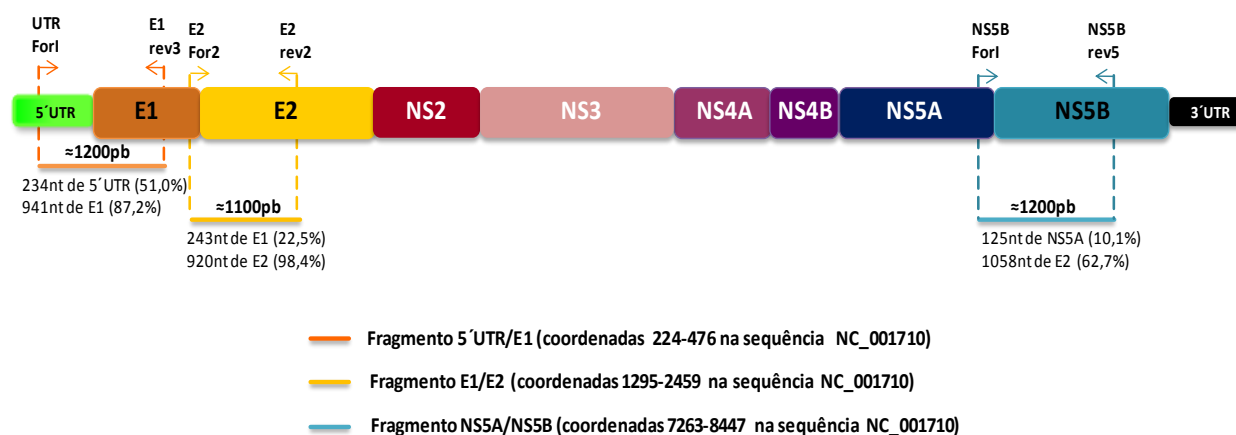


Figura 3.9. Representação esquemática do genoma de GBV-C. O tamanho dos produtos obtidos nas segundas reacções de PCR para as regiões estudadas (5' UTR/E1, E1/E2 e NS5A/NS5B) e a correspondência em percentagem da fracção de cada região amplificada relativa a cada fragmento, encontram-se indicados.

Em ambas as situações foi necessário um ajuste no protocolo de amplificação, nomeadamente no que diz respeito às temperaturas de hibridação utilizadas (de 55°C para 58°C) nos últimos 10 ciclos de amplificação. Esta alteração foi justificada pela tendência de formação de fragmentos inespecíficos, detectados através da análise de alíquotas das reacções de amplificação em gel de

agarose. Contudo, apesar da optimização das condições de amplificação, a presença sistemática de alguns dos produtos, amplificados de forma não específica, manteve-se em determinados casos, ainda que tenham sido testados gradientes de temperatura de hibridação (55°C-65°C), diferentes quantidades de DNA matriz (1-4µl) e diluições (1/20 e 1/50) do produto obtido na primeira reacção de PCR.

Relativamente à região NS5A/NS5B, não foi possível a obtenção de um produto de amplificação para 4 das amostras biológicas em análise (3638, 6004, 7888 e 8731). Nestes casos, recorreu-se a uma alteração das temperaturas de hibridação dos *primers*, através de um gradiente onde ocorreu uma variação entre 45°C e 58°C na primeira reacção de PCR e entre 50°C e 58°C na segunda. De igual forma, a quantidade de DNA matriz foi aumentada até 5µl na passagem da primeira para a segunda reacção de PCR, não tendo sido, no entanto, detectada como resultado de nenhuma das condições testadas, a presença de quaisquer produtos.

Para as amostras onde se verificou a amplificação do produto desejado, este revelou-se sob a forma de um amplicão específico com o tamanho esperado ($\approx 1,1$ kpb para a região E1/E2, Figura 3.10 A e $\approx 1,2$ kpb correspondente à região NS5A/NS5B, Figura 3.10 B), acompanhado por vezes de fragmentos inespecíficos (Figura 3.10 B).

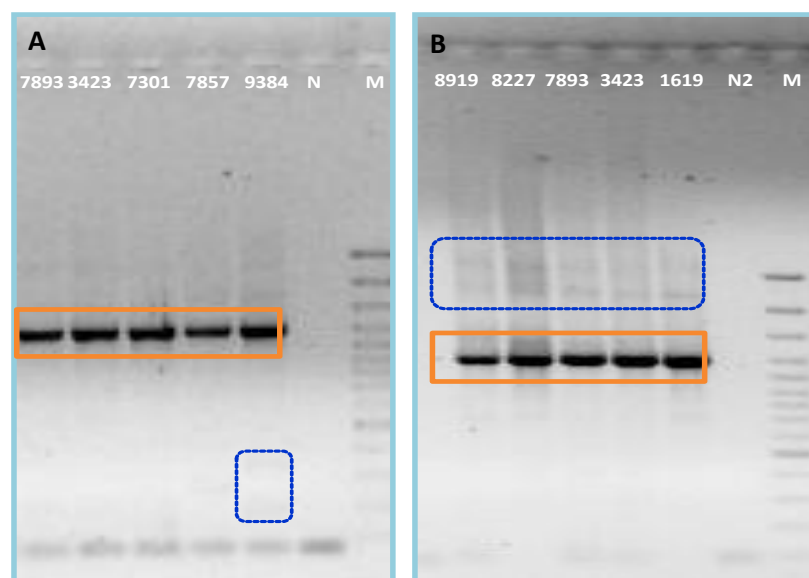


Figura 3.10. Observação, em gel de agarose, dos produtos referentes à amplificação, por *nested* PCR, da região E1/E2 **A**) e NS5A/NS5B, **B**). As pistas estão identificadas com o número correspondente à amostra em estudo; M - marcador de massa molecular (*GeneRuler™100bp DNA Ladder plus*); N - controlo negativo. Os fragmentos correspondentes ao produto de amplificação pretendido, encontram-se indicados por um rectângulo laranja e a presença de produtos inespecíficos demarcada com pontilhado azul.

A importância destes fragmentos nas misturas de PCR foi considerada irrelevante devido à sua quantidade diminuta, comparativamente com o fragmento pretendido. Tal facto veio a confirmar-se para a maioria dos casos, por exemplo, através da obtenção de resultados de sequenciação não ambíguos.

3.3. Clonagem molecular de sequências genómicas de GBV-C em vectores plasmídicos

Após amplificação dos diferentes fragmentos de DNA pretendidos, foi obtida a sua respectiva sequência nucleotídica. De um total de 86 amostras analisadas, a sequenciação directa dos amplicões purificados, correspondentes a 8 fragmentos amplificados cobrindo a região NS5A/NS5B (19,51%) e 2 fragmentos amplificados da região E1/E2 (4,44%), revelou resultados considerados de qualidade insatisfatória, traduzidos pela baixa qualidade das sequências obtidas (cromatogramas de difícil interpretação/ambíguos). Para estes casos (Tabela 3.2), efectuou-se a clonagem dos fragmentos em vectores plasmídicos (pGEM[®]-T Easy ou pGEM[®]-T, de acordo com a disponibilidade existente no laboratório) usando *E. coli* JM109 como hospedeiro de clonagem.

Tabela 3.2. Identificação dos fragmentos clonados, com referência à respectiva região genómica, vector plasmídico e enzima(s) de restrição utilizada(s) na sua análise.

Amostras analisadas	Região do genoma	Vector plasmídico	Enzima(s) de Restrição
94	NS5A/NS5B	pGEM [®] -T Easy	EcoRI
94	E1/E2	pGEM [®] -T Easy	EcoRI
111	NS5A/NS5B	pGEM [®] -T	NcoI/Sall
1642	5' UTR	pGEM [®] -T	NcoI/Sall
3013	NS5A/NS5B	pGEM [®] -T Easy	EcoRI
6073	NS5A/NS5B	pGEM [®] -T Easy	EcoRI
7251	NS5A/NS5B	pGEM [®] -T Easy	EcoRI
7251	E1/E2	pGEM [®] -T Easy	EcoRI
7857	NS5A/NS5B	pGEM [®] -T	NcoI/Sall
8919	NS5A/NS5B	pGEM [®] -T	NcoI/Sall
9709	NS5A/NS5B	pGEM [®] -T	NcoI/Sall

Os sistemas de clonagem utilizados permitiram a selecção, *a priori*, dos clones bacterianos contendo plasmídios recombinantes pela observação de coloração azul (clones sem o vector recombinante) ou branca (clones com vector recombinante) nas colónias bacterianas crescidas em meio selectivo, como resultado do fenómeno de α -complementação. Após expansão dos potenciais clones recombinantes em meio adequado, e subsequente extracção do respectivo DNA plasmídico, as moléculas de DNA recombinante foram identificadas pela análise do seu padrão electroforético em gel de agarose (a título exemplificativo, ver Figura 3.11 A). Esta análise foi baseada na comparação da migração de alíquotas do extracto de DNA plasmídico, proveniente de diferentes clones, relativamente a um extracto plasmídico contendo moléculas de referência de tamanho semelhante ao vector utilizado. As potenciais moléculas recombinantes deveriam apresentar uma migração retardada relativamente às de referência (Figura 3.11), como resultado da inserção de fragmentos de DNA exógeno (insertos), e do conseqüente aumento da respectiva massa molecular, quando comparada com o vector de clonagem.

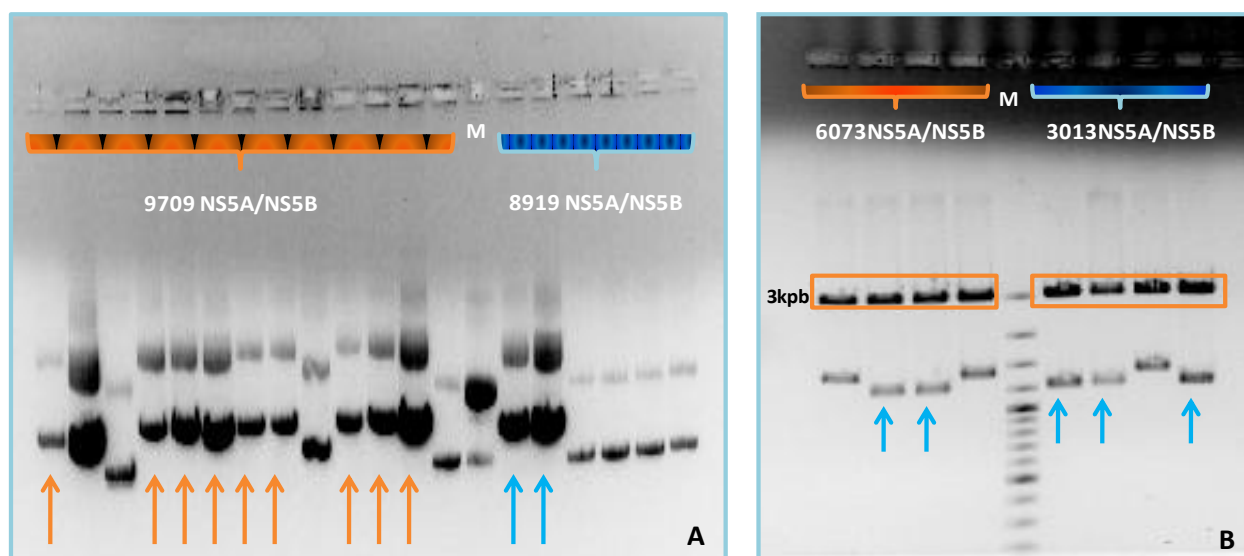


Figura 3.11. A) Análise do padrão de migração, em gel de agarose, do DNA plasmídico de alguns dos clones bacterianos obtidos na sequência das tentativas de clonagem de fragmentos de DNA referentes à região NS5A/NS5B (correspondentes às amostras 9709 e 8919). As setas indicam potenciais moléculas de DNA plasmídico recombinante. M - marcador de massa molecular (plasmídio de referência com 3kbp); B) Perfil de digestão resultante da hidrólise, pela enzima EcoRI, de potenciais plasmídios recombinantes contendo amplicões NS5A/ NS5B obtidos para as amostras 6073 e 3013. As bandas respectivas ao vector plasmídico pGEM[®]-T Easy linearizado (com 3kbp) estão indicadas por um rectângulo laranja. As setas assinalam a presença de um fragmento com o tamanho correspondente ao DNA de interesse (inserto), $\approx 1,2$ kbp. M - marcador de tamanho molecular (*GeneRuler[™] 100bp DNA Ladder plus*).

A subsequente caracterização dos clones de DNA recombinante, desta forma identificados, baseou-se na análise dos perfis electroforéticos dos fragmentos obtidos mediante a hidrólise dos plasmídeos em estudo com diferentes enzimas de restrição, cujos locais de reconhecimento e corte flanqueiam o local de clonagem múltipla dos vectores utilizados (Tabela 3.2). A confirmação da presença do fragmento do DNA de interesse foi indicada pelo perfil de digestão enzimática obtido usando a enzima EcoRI. Neste, foi revelada, como esperado, a presença de uma banda de 3kpb, correspondente ao vector plasmídico linearizado, e outra de menores dimensões, respectiva ao fragmento de DNA clonado ($\approx 1,2$ kpb no caso de um fragmento da região NS5A/NS5B – Figura 3.11 B - e com $\approx 1,1$ kpb para a região E1/E2). Por outro lado, o padrão de restrição obtido mediante a digestão das moléculas de DNA extraídas simultaneamente com as enzimas NcoI e SalI, revelou a presença de três fragmentos de DNA, um com 3kpb relativo ao vector pGEM[®]-T linearizado e dois com 500pb e 1000pb, relativos ao fragmento de DNA clonado.

O DNA plasmídico contido nos clones recombinantes de interesse (para cada amostra em estudo) foi extraído em pequena escala e, em seguida, sequenciado com os *primers* universais T7 (directo) e Sp6 (reverso). As reacções de sequenciação efectuadas, num total de 10, resultaram em cromatogramas de elevada qualidade (Figura 3.12, cromatograma inferior), especialmente quando comparados aos obtidos como consequência da sequenciação directa dos produtos de PCR inicialmente obtidos (Figura 3.12, cromatograma superior).

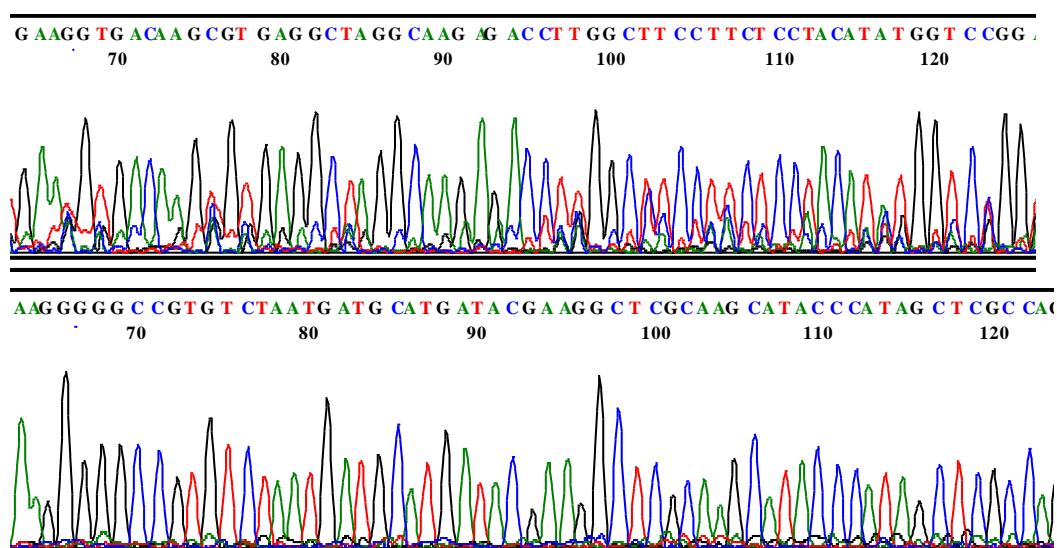


Figura 3.12. Exemplo de cromatogramas relativos às sequências nucleotídicas obtidas por sequenciação directa do produto de amplificação NS5A/NS5B referente à amostra 8919 (topo). O cromatograma em baixo espelha o melhoramento na qualidade das sequências nucleotídicas obtidas para a mesma amostra antes e depois do processo de clonagem do fragmento original.

3.4. Análise de sequências nucleotídicas

3.4.1. Reconstrução de filogenias baseadas na análise de sequências de GBV-C

A caracterização genética de estirpes de GBV-C circulantes na população estudada foi realizada com base na análise de 172 sequências nucleotídicas. Estas foram em número de 82 para a região NS5A/NS5B (n=41 amostras) e 90 para a região E1/E2 (n=45 amostras), correspondendo a pares de sequências sobreponíveis, obtidas por utilização de *primers* que permitiram a sequenciação, em sentidos opostos, dos fragmentos em análise. As referidas sequências serviram de base à construção de igual número de sequências contíguas (*contigs*), com base na análise dos pares de sequências obtidas para cada região/amostra analisada. Depois de corrigidas manualmente, efectuou-se uma análise bioinformática com as ferramentas BLASTn e BLASTx (acessíveis em <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) para confirmação da sua identidade. As sequências em questão possibilitaram, posteriormente, a construção, em separado, de alinhamentos múltiplos, um para a região E1/E2 e outro para a região NS5A/NS5B. Estes incluíram, não só as sequências nucleotídicas obtidas no decurso deste estudo, como também 38 sequências usadas como referência, correspondentes ao genoma completo de GBV-C, e que contemplaram representantes dos 6 genótipos de GBV-C reconhecidos, tal como definidos por Muerhoff e colaboradores (Muerhoff *et al.*, 2006). De entre estas, deverá salientar-se que 3 foram apontadas como recombinantes (AB013501 - Konomi *et al.*, 1999; AB0021287 - Naito *et al.*, 1999; U75356.1 - An *et al.*, 2000). Foram igualmente utilizadas 54 sequências de referência correspondentes a genomas parciais, de entre as quais, 44 são respeitantes à região E2 e 10 relativas à região NS5B (números de acesso no Anexo V). Tendo por suporte estes dois alinhamentos, uma vez que só assim se torna possível a identificação de posições (nucleótidos) homólogas (homólogos posicionais), foram traçadas árvores filogenéticas relativas às duas regiões genómicas em estudo. Para caracterização genética das estirpes de GBV-C, recorreu-se a três métodos de reconstrução filogenética: utilização de algoritmos de adição como o método de junção-de-vizinhos (*Neighbor-Joining* - NJ), análise Bayesiana com base no processo aleatório de cadeias de Monte-Carlo (*Markov Chain Monte Carlo* – MCMC) e pesquisa de árvores filogenéticas por aplicação do critério de Máxima Verosimilhança (*Maximum Likelihood* – ML). O objectivo final foi a construção de árvores filogenéticas que fossem representações, tão fiéis quanto possível, das relações de ancestralidade existentes entre os *taxa* em estudo.

3.4.2. Relações filogenéticas entre estirpes de GBV-C estimadas pela análise da região E1/E2

A análise filogenética das sequências nucleotídicas correspondentes à região E1/E2 compreendeu o conjunto de sequências (descrito em 3.4.1) com o qual foram delineados dois alinhamentos múltiplos distintos. O primeiro, construído com base em 127 sequências de 760nt alinhadas após edição, incluiu 82 sequências de referência e 45 sequências obtidas no decurso deste estudo. Durante o processo de edição manual foram desprezadas 4 sequências de referência (AY951953, AY951965, AY951979, AY951978, correspondentes ao genótipo 5) uma vez que se verificou que, desta forma, se poderiam obter alinhamentos num total de 123 sequências de maior dimensão (960nt). As árvores obtidas com ambos os alinhamentos foram construídas com base na aplicação dos três métodos de inferência filogenética já mencionados (secção 2.11.2). No entanto, a representação gráfica daquelas que resultaram do alinhamento com menor número de sequências (Figura 3.14 A), foi restringida ao método de junção-de-vizinhos, onde se denotou uma dispersão praticamente idêntica das sequências estudadas, comparativamente à que resultou da análise do alinhamento de 127 sequências (Figura 3.14 B). O mesmo sucedeu relativamente aos restantes métodos utilizados, não existindo diferenças significativas entre as topologias das árvores em que foram analisadas um maior número de sequências parciais das obtidas com base num menor conjunto de sequências com maior tamanho (960nt). Por esta razão, as segundas, não se encontram representadas neste manuscrito.

Tanto as árvores filogenéticas que tiveram por base o método de NJ, onde foram considerados significativos os agrupamentos que apresentaram valores de *bootstrap* iguais ou superiores a 75%, como aquelas suportadas por valores de probabilidade *a posteriori* (>0,90) (análise Bayesiana), como ainda as estabelecidas pelo critério de máxima verosimilhança (ML), apresentaram topologias semelhantes. Particularmente, verificou-se uma distribuição das 45 estirpes de GBV-C em estudo em agrupamentos monofiléticos associados a genótipos específicos. A segregação obtida revelou 10 das estirpes de GBV-C analisadas como pertencentes ao genótipo 1 e 34 ao genótipo 2 (G1 e G2_{Total}, respectivamente, Figura. 3.13). Ainda dentro do genótipo 2, das 34 sequências neste agrupadas, 9 projectam-se num grupo afastado das demais, identificado como G2^a (Figuras 3.14 e 3.15). No entanto esta divergência foi apenas suportada (com um valor de probabilidade *a posteriori* de 1.00) pela análise Bayesiana efectuada.

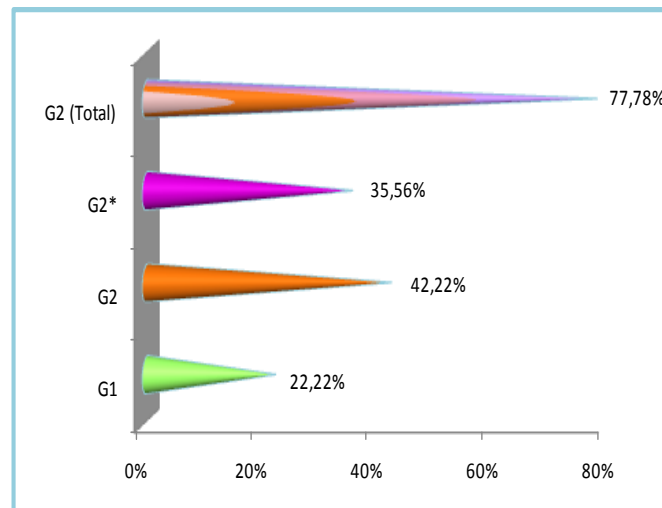
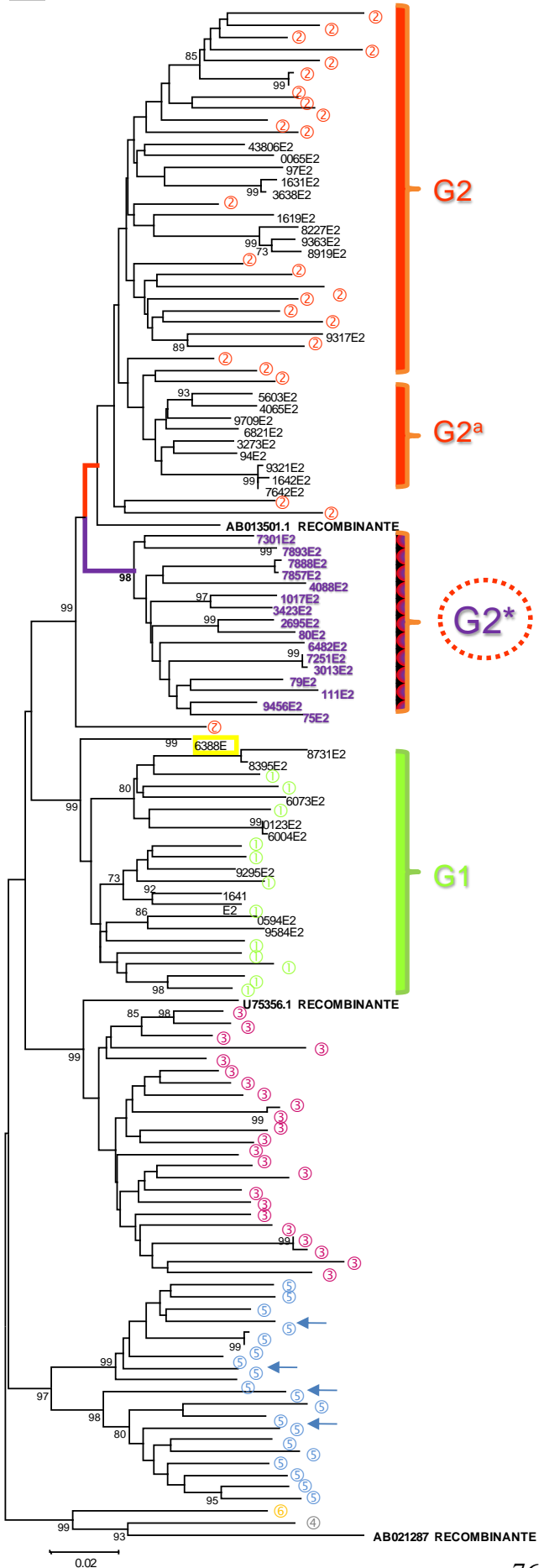


Figura 3.13. Distribuição (percentagem) das estirpes de GBV-C em estudo, agrupadas em *clusters* monofiléticos pela análise da região E1/E2. Gen2 (Total) indica a totalidade de sequências que foram agrupadas neste genótipo (conjunto das sequências em estudo associadas a Gen2, Gen2^a e G2*).

B



C

G1 - AB013500, AB003291, AY951958, AY951961, AY951969, AY951970, AY951972, AY951973, AY951975, AY951980, U36380, U87659 (①)

G2 - AB003289, AB021287, AF031829, AF081782, AF104403, AF121950, AF309966, AY951954; AY951974, AY196904, D87255, D90600, NC_001710, U45966, U87653, U87654, U87655, U87656, U87657, U87660, U87661, U87662, U87663, U87664 (②)

G3 - AB000150, AB000158, AB000164, AB003288, AB003290, AB003293, AB008335, AB008342, AF006500, AF017532, AF017533, D87263, D87708, D87709, D87710, D87711, D87712, D87713, D87714, D87715, D90601, U94695 (③)

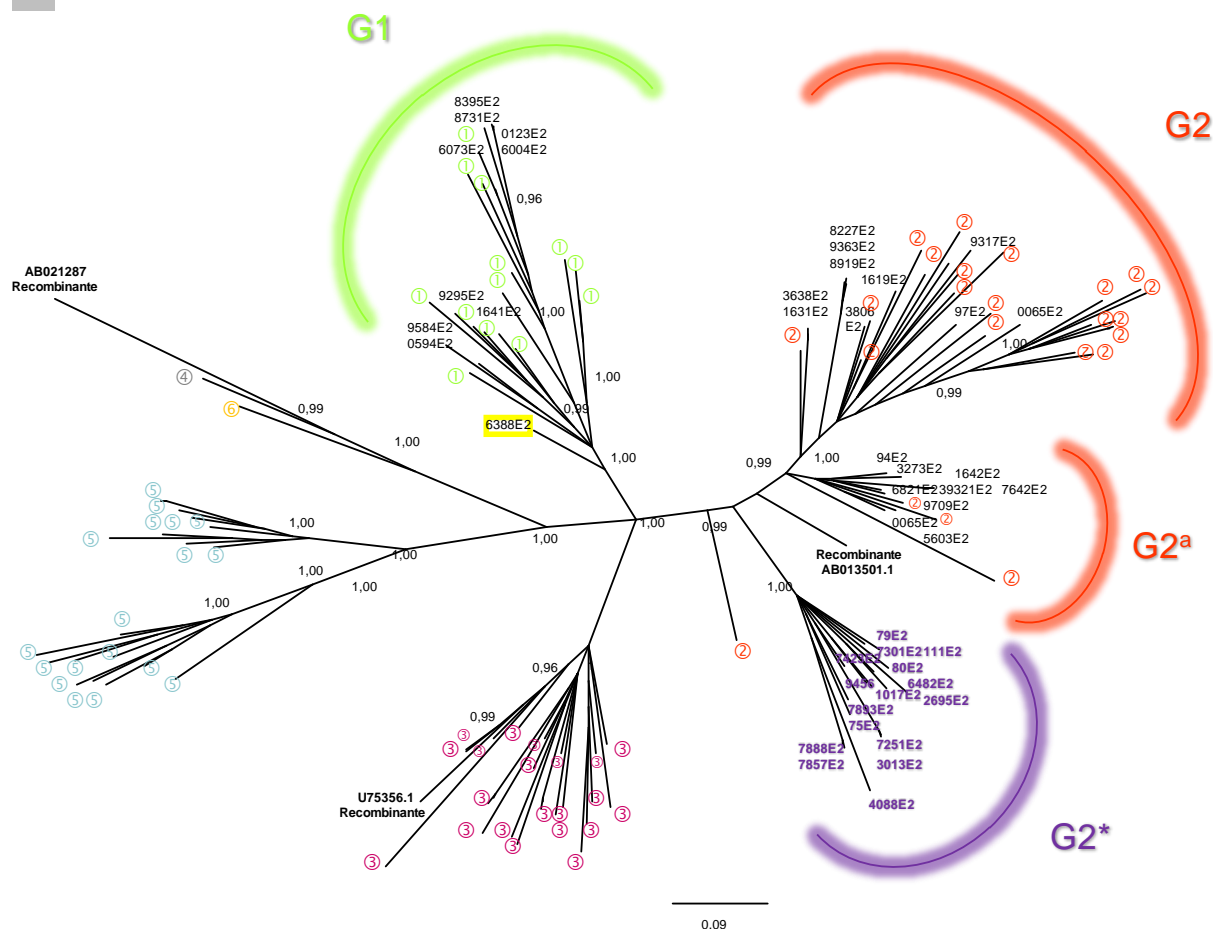
G4 - AB018667 (④)

G5 - AY949771, **AY951953**, AY951955, AY951956, AY951957, AY951959, AY951960, AY951962, AY951963, AY951964, **AY951965**, AY951966, AY951967, AY951968, AY951971, AY951976, AY951977, **AY951978**, **AY951979** (⑤)

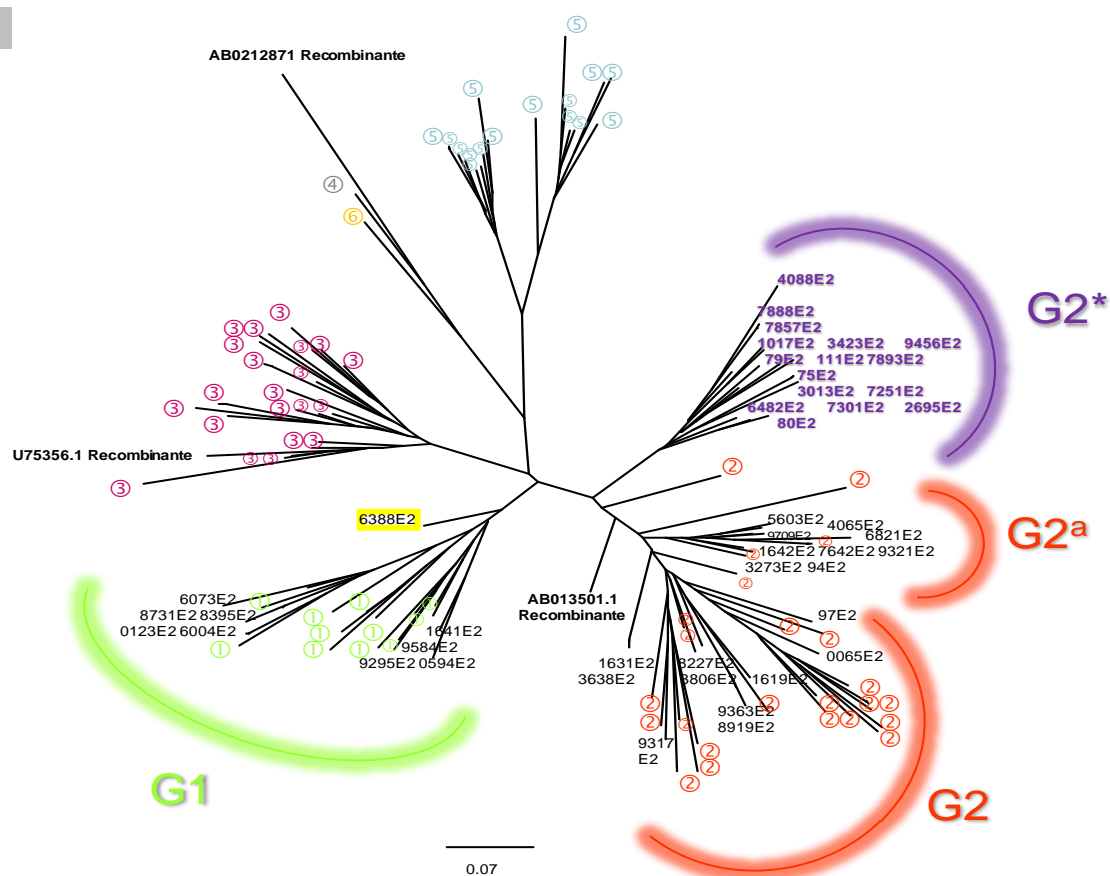
G6 - AB003292 (⑥)

Figura 3.14. Relações filogenéticas entre as estirpes de GBV-C em estudo, baseadas na análise das sequências nucleotídicas da região E1/E2. As árvores filogenéticas representadas, ambas sem raiz, foram obtidas pelo método NJ, a partir de matrizes de distâncias calculadas pelo método K2P. Estão também indicados os valores de *bootstrap* (percentagem de 1000 réplicas) considerados significativos. A escala representa 2% de distância genética. As sequências das amostras estudadas estão indicadas pela respectiva numeração e com o sufixo E2. As sequências de referência utilizadas, por questões gráficas e de forma a facilitar a interpretação das árvores obtidas, estão evidenciadas pelo número do genótipo correspondente e assinaladas com diferentes cores. **A)** Árvore filogenética obtida com 78 sequências de referência juntamente com as 45 estirpes de GBV-C analisadas, perfazendo uma total de 123 sequências nucleotídicas. **B)** Árvore filogenética obtida com 82 sequências de referência e 45 estirpes de GBV-C em estudo, perfazendo uma análise de 127 sequências nucleotídicas. As setas azuis indicam as 4 sequências de referência extra utilizadas na construção da referida árvore. O número de acesso respectivo a estas sequências está referido em **C**, com fonte azul. **C)** Correspondência entre as sequências de referência utilizadas com o respectivo genótipo.

A



B



C

G1 - AB013500, AB003291, AY951958, AY951961, AY951969, AY951970, AY951972, AY951973, AY951975, AY951980, U36380, U87659 (①)
G2 - AB003289, AB021287, AF031829, AF081782, AF104403, AF121950, AF309966, AY951954; AY951974, AY196904, D87255, D90600, NC_001710, U45966, U87653, U87654, U87655, U87656, U87657, U87660, U87661, U87662, U87663, U87664 (②)
G3 - AB000150, AB000158, AB000164, AB003288, AB003290, AB003293, AB008335, AB008342, AF006500, AF017532, AF017533, D87263, D87708, D87709, D87710, D87711, D87712, D87713, D87714, D87715, D90601, U94695 (③)
G4 - AB018667 (④)
G5 - AY949771, AY951955, AY951956, AY951957, AY951959, AY951960, AY951962, AY951963, AY951964, AY951966, AY951967, AY951968, AY951971, AY951976, AY951977 (⑤)
G6 - AB003292 (⑥)

Figura 3.15. Relações filogenéticas entre as 45 estirpes de GBV-C em estudo, baseadas na análise das sequências nucleotídicas da região E1/E2, e 78 sequências de referência relativas aos 6 genótipos reconhecidos de GBV-C. As sequências das amostras estudadas estão identificadas pela numeração referente às amostras de onde foram amplificadas, seguidas do sufixo E2. As sequências de referência utilizadas, por questões gráficas, estão evidenciadas pelo número do genótipo correspondente (com diferentes cores). As árvores filogenéticas radiais representadas foram obtidas por 2 métodos de reconstrução distintos. **A)** Inferência filogenética baseada numa análise Bayesiana ao longo de 6×10^6 gerações, tendo por base o modelo GTR. Foram considerados significativos os agrupamentos que apresentam valores de probabilidade *a posteriori* $>0,9$. **B)** Árvore filogenética obtida com base no critério de máxima verosimilhança. A proporção de sítios invariáveis foi equivalente a 0,44 e o parâmetro $\alpha=0,67$. **C)** Correspondência entre as sequências de referência utilizadas com o respectivo genótipo.

Um conjunto de 16 estirpes de GBV-C, também elas incluídas no agrupamento formado pelas sequências de referência classificadas como pertencentes ao genótipo 2, revelaram divergir das demais, formando um sub-grupo distinto, o qual passou a ser denominado por G2*. Nas árvores obtidas, G2* apresentou-se suportado por um elevado valor de *bootstrap* (98%) na árvores de NJ, bem como pelo máximo de probabilidade *a posteriori* (1,00) detectada com a análise Bayesiana efectuada (Figuras 3.14 e 3.15 A, respectivamente). O agrupamento G2 (genótipo 2) revelou-se, por isso, como um grupo caracterizado por uma elevada diversidade genética, traduzida pela formação de, pelo menos, 3 sub-grupos aparentemente distintos. Dois destes, G2* e o agrupamento que contém as demais sequências G2 (excluindo G2^a), foram revelados pelos três métodos de reconstrução filogenética utilizados.

A existência do sub-grupo G2* foi confirmada com recurso à ferramenta PCOORD, que tem por base a procura e associação de padrões significativos entre as sequências em análise. Nesse sentido construiu-se um alinhamento múltiplo que incluiu as 45 sequências nucleotídicas amplificadas de GBV-C (respeitantes à região E1/E2). Os valores obtidos por esta análise registaram 25,90% de variabilidade entre o total das sequências analisadas. A representação

gráfica deste resultado indica uma clara distinção entre os grupos de sequências correspondentes ao genótipo 1, 2 e ao sub-grupo G2* (Figura 3.16).

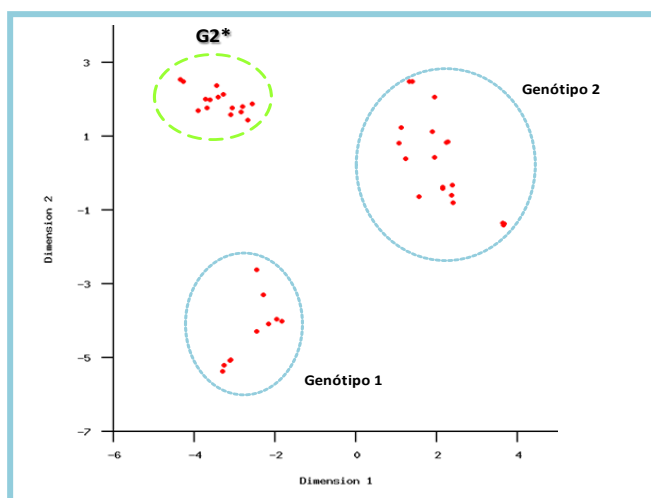


Figura 3.16. Representação gráfica da análise obtida pela ferramenta PCOORD, para os grupos correspondentes ao genótipo 1, 2 e G2*. Os eixos representam as duas dimensões que foram primeiramente extraídas e caracterizam 25,90% das diferenças encontradas entre os 3 grupos.

Ainda no que diz respeito à análise das árvores filogenéticas, verificou-se que as sequências 1631E2 e 3638E2 (genótipo 2), 9321E2, 1642E2, e 7642E2 (incluídas em G2^a), 7888E2, 7857E2, 7251E2 e 3013E2 abrangidas por G2* e 0123E2 e 6004E2 (genótipo 1), segregavam em agrupamentos monofiléticos suportados pela análise de *bootstrap* efectuada com valores elevados (100%, Figura 3.14 A e 99%, Figura 3.14 B) e um valor de probabilidade *a posteriori* de 1,00 na análise Bayesiana (Figura 3.15 A), sugerindo a existência de uma estreita relação epidemiológica entre elas.

Nas quatro árvores construídas foi igualmente evidente a separação das sequências recombinantes (incluídas neste estudo como amostras de referência), as quais ocuparam posições divergentes relativamente aos ancestrais comuns dos agrupamentos específicos para cada genótipo respectivo. Foi disto excepção a posição da sequência U75356, a qual, nas árvores Bayesiana e de ML, fez parte do agrupamento definido pelas referências correspondentes ao genótipo 3. Da mesma forma, a estirpe 6388E2 (destacada por um rectângulo amarelo nas Figuras 3.14 e 3.15) segregou prematuramente das demais incluídas no grupo formado pelo genótipo 1 (G1, sendo este relativamente homogêneo), pressupondo uma natureza recombinante, ou uma sequência altamente divergente das demais representantes de G1.

A hipótese de se tratar de uma sequência recombinante foi testada por uma análise de *bootscanning* (onde foram considerados significativos os valores acima de 70%) a qual evidenciou fenômenos de recombinação intragenotípica com o genótipo 1 na região central da sequência em estudo (Figura 3.17). As duas sequências de referência de genótipo 1 que revelaram maior semelhança com a amostra estudada foram as identificadas com os números de acesso AB003291 e U36380. No entanto, apesar da evidência de prováveis eventos recombinantes intragenotípos, não foi possível determinar a origem genética das extremidades 3' e 5' por ausência de homologia com as sequências de referência (Figura 3.17).

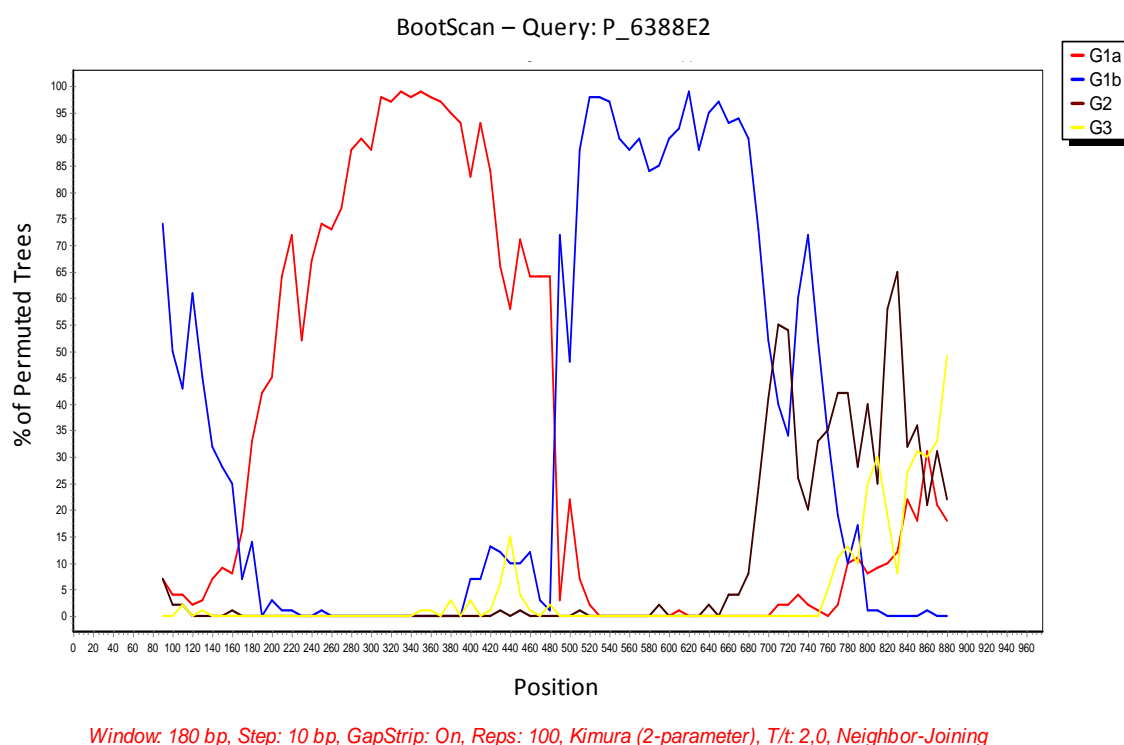


Figura 3.17. Análise de *bootscanning* por *SimPlot* (janela de 180pb, passo de 10pb e 100replicados; inferência filogenética com base no método de NJ) da sequência 6388E2. As letras da legenda dos gráficos G2 e G3 fazem correspondência com os genótipos 2 e 3, respectivamente. G1a e G1b correspondem às sequências AB003291 e U36380, respectivamente.

3.4.2.1. Caracterização genética do agrupamento G2*

De forma a determinar se o conjunto de sequências que agrupadas em G2* eram suficientemente divergentes do grupo monofilético mais próximo (formado pelas demais sequências do genótipo 2) para comporem um genótipo distinto, procedeu-se à análise de distâncias genéticas. Esta envolveu a comparação das distâncias calculadas quando sequências pertencentes a um mesmo

genótipo (intra-genótipo), ou a genótipos distintos (inter-genótipos) foram comparadas entre si. Este cálculo envolveu não só as sequências de GBV-C obtidas neste estudo, como também as de referência que representam os genótipos 1, 2, 3 e 5 (os genótipos 4 e 6, devido à reduzida quantidade de sequências disponíveis, não foram incluídos nesta análise).

Os valores obtidos, correspondendo a distâncias corrigidas pela fórmula de K2P, foram agrupados em classes e a distribuição dos resultados auferidos é a apresentada graficamente na Figura 3.18 A. Pela análise do histograma traçado (Figura 3.18 A), verifica-se uma distribuição das distâncias genéticas relativas a G2* (barras rosa), exactamente na região central entre a distribuição de distâncias intragenóticas e intergenóticas. Assim, a comparação das distâncias genéticas calculadas, não permite apoiar, de forma indubitável, a hipótese de que G2* corresponda a um novo genótipo de GBV-C, apesar da topologia das árvores filogenéticas obtidas com os três métodos de construção sugerir, claramente, uma divergência do grupo G2* (caracterizado por uma diversidade genética de 11,80%) relativamente às restantes sequências incluídas no genótipo 2, a qual foi apoiada pela análise PCOORD. Consequentemente, G2* poderá representar um sub-tipo específico, presente na população portuguesa residente na área metropolitana de Lisboa, e, como tal confinado a uma zona geográfica singular.

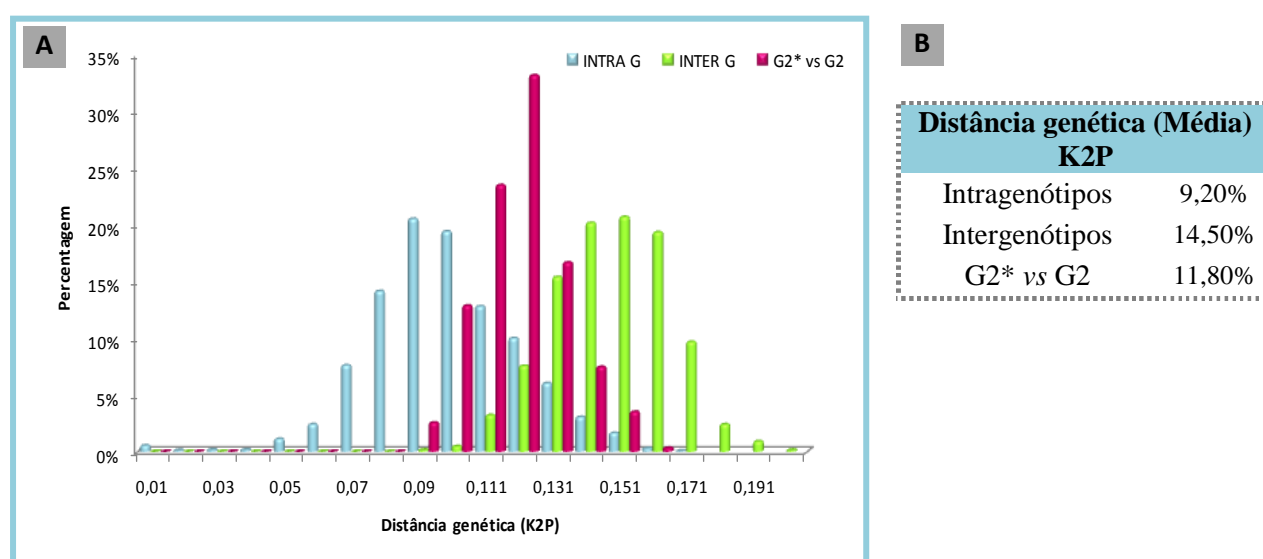


Figura 3.18. A) Representação gráfica das distâncias genéticas inter e intragenótipos, considerando os genótipos 1,2,3 e 5 de GBV-C e o grupo G2* vs G2. **B)** Tabela com o cálculo das distâncias genéticas médias intra e intergenótipos obtidas pelo programa Mega 4.0.2. com base no método de Kimura a dois parâmetros.

Com base no mesmo conjunto de sequências, as distâncias intra e intergenótipos entre os vários grupos em análise foram comparadas com aquela revelada pelo grupo intraG2*. Pela observação

da Tabela B (Figura 3.19), o agrupamento G5 é o mais heterogêneo e de modo inverso, o G2*, aquele que representa maior homogeneidade a nível das sequências analisadas. Os grupos G3 e G5 são aqueles que se encontram mais afastados filogeneticamente (15,80%), e tal como previsto, os grupos de sequências caracterizados como G2* e G2 são os mais próximos entre si (11,80%). O histograma que caracterizou a homogeneidade do grupo G2*, através do cálculo da distância intraG2*, encontra-se representado na Figura 3.19 A. Foi evidente uma distribuição deste grupo inserida no agrupamento “intra G”, revelando a consistência genética destas sequências.

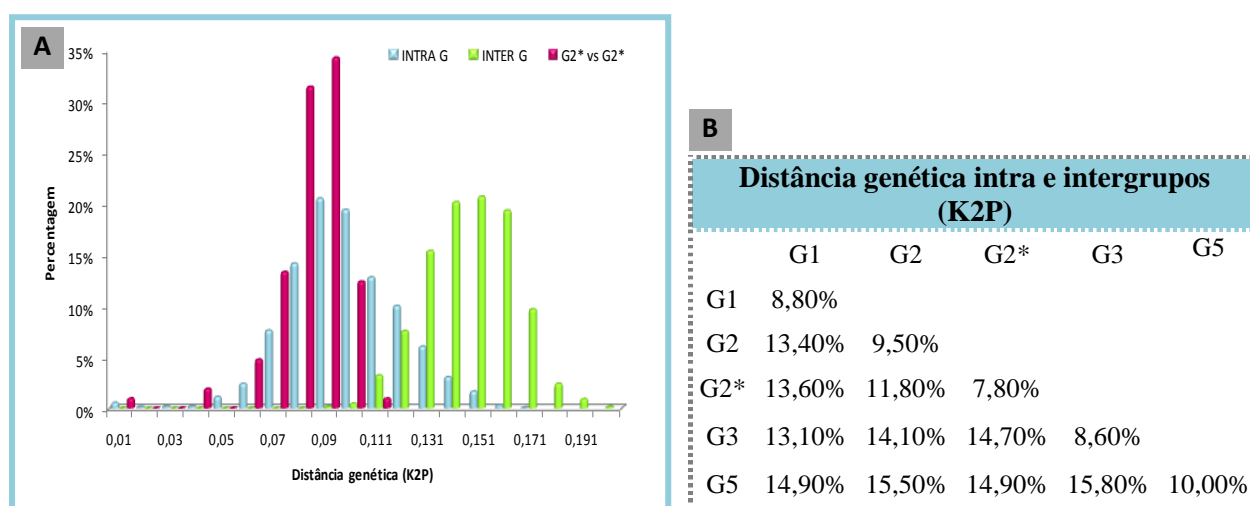


Figura 3.19. A) Representação gráfica das distâncias genéticas inter e intragrupos, considerando os genótipos G1, G2, G3, G5 e o grupo G2*. B) Tabela com indicação das distâncias genéticas calculadas dentro, e, entre os respectivos grupo/genótipos indicados (obtidas pelo programa Mega 4.0.2. com base no método de Kimura a dois parâmetros).

A ocorrência de substituições nucleotídicas sinónimas (d_S) e não sinónimas (d_N) foi estimada através do cálculo da razão d_S/d_N , a partir de um alinhamento múltiplo de sequências nucleotídicas, alinhadas por codão. Este tipo de análise permite-nos inferir sobre o tipo de pressão selectiva que opera sobre as sequências estudadas. Nesta, foram incluídos 2 conjuntos de sequências: 1) o grupo de 45 sequências E1/E2 de GBV-C obtidas no decurso deste trabalho 2) as sequências em estudo que formam parte do genótipo 2 (G2*+G2^a+G2), perfazendo um total de 35 sequências. Os 2 alinhamentos resultantes foram analisados com a ferramenta SNAP de acordo com o método de Nei e Gojobori. Este cálculo permite avaliar se a variabilidade genética destes agrupamentos se traduz em variações na sequência primária das proteínas ou se, por outro lado, as alterações registadas quando as sequências nucleotídicas são comparadas, são alterações de tipo silencioso. A razão d_S/d_N foi calculada para as sequências nucleotídicas referidas, e a

média de valores obtida foi de 25,8 para o primeiro conjunto de sequências analisadas, e de 0,92 para as 35 sequências agrupadas no genótipo 2. No primeiro caso, os valores auferidos são muito superiores a 0, o que indica que as sequências analisadas estão sobre uma aparente pressão negativa (tendência para a conservação a nível de aminoácidos pela ocorrência de substituições silenciosas dos mesmos), facto concordante com a distribuição cumulativa dos valores de d_S e d_N ao longo das sequências comparadas codão por codão (Figura 3.20 A). Em termos evolutivos, e no que respeita às sequências analisadas de genótipo 2, verifica-se uma maior tendência para a ocorrência de substituições não sinónimas (Figura 3.20 B), também evidenciada a nível individual de d_S/d_N por codão (grande maioria com d_S/d_N próximos do valor médio calculado).

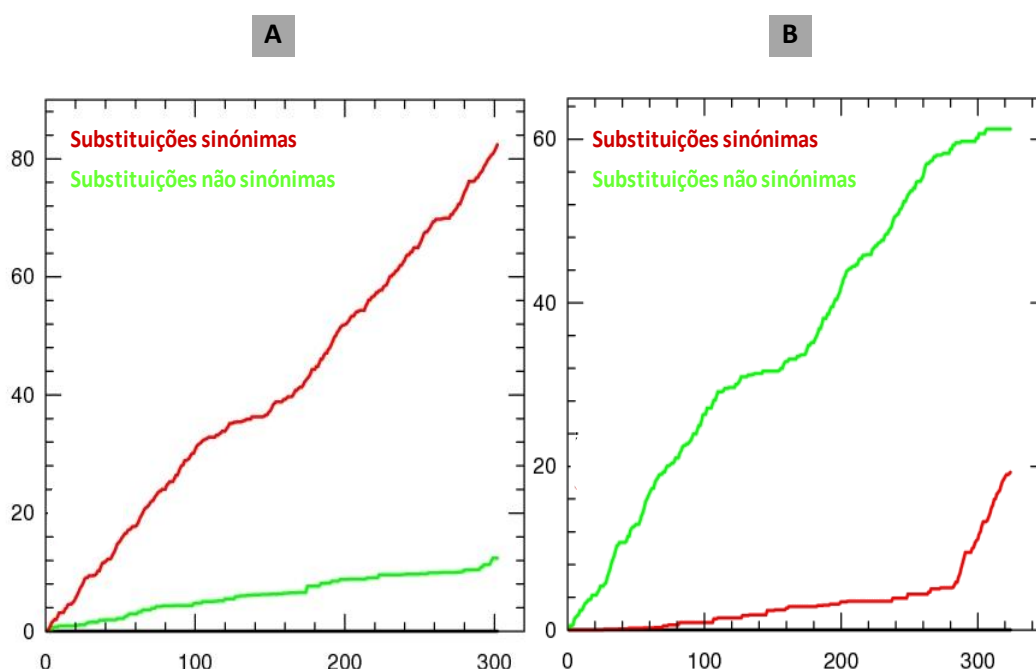


Figura 3.20. Representação gráfica do efeito acumulativo da ocorrência de substituições sinónimas e não sinónimas por codão. Em **A)** estão representados os resultados referentes ao conjunto de 45 sequências de GBV-C analisadas e em **B)**, a representação gráfica obtida para o conjunto de sequências alusivas ao genótipo 2.

No sentido de averiguar o grau de variabilidade relativa às posições de aminoácidos, um grupo de sequências nucleotídicas foi traduzido em resíduos de aminoácidos e editado, de forma a construir alinhamentos múltiplos de sequências proteicas. A partir desses alinhamentos foi possível avaliar a conservação das posições aminoacídicas com base no cálculo da Entropia de Shannon (usando a ferramenta *Entropy-2*, acompanhada de uma análise estatística de 1000 repetições com reposição, definindo o valor de $P=5/1000$).

Como grupo alvo (*query*) foram estudadas as sequências pertencentes a G2*, uma vez que segregou de forma particular do genótipo 2, e definido como suporte de comparação

(*background*), o conjunto das 18 sequências de GBV-C em estudo, que segregaram com o genótipo 2, juntamente com 19 sequências de referência G2. A comparação da entropia entre os dois grupos em análise (*query vs background*), demonstrou um elevado grau de conservação da posição de determinados aminoácidos relativamente ao grupo G2* com valores estatisticamente significativos ($P < 0,005$), nomeadamente nos resíduos A58, S110, Q134 e F182 (Figura 3.21). Por sua vez, o resíduo N86 (Figura 3.21) encontra-se mais conservado no grupo de sequências que constituem o respectivo *background* (G2).

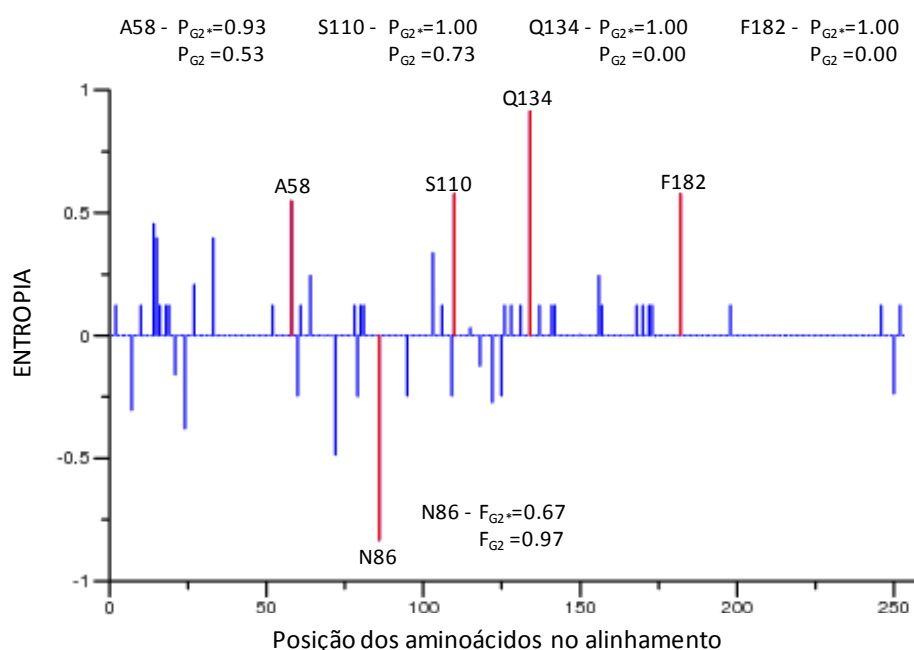


Figura 3.21. Representação gráfica das posições conservadas dos resíduos de aminoácidos, assinaladas por barras vermelhas, com indicação dos valores de probabilidade a estes associada. Comparação das posições de aminoácidos entre o grupo de sequências pertencentes a G2* e o conjunto de sequências de GBV-C em estudo alusivas ao genótipo 2.

O estudo da eventual dependência entre as variáveis de carácter epidemiológico das amostras em estudo e a sua associação a determinado genótipo, através do teste de χ^2 (com correcção de *Yates*), revelou uma associação estatisticamente significativa entre o genótipo 1 e etnia Africana (respeitante à nacionalidade dos indivíduos em amostragem), com $P = 0,005$. Para os restantes casos a análise estatística mostrou uma independência das variáveis estudadas ($P > 0,05$).

3.4.3. Relações filogenéticas entre estirpes de GBV-C estimadas pela análise da região NS5A/NS5B

Em paralelo com o que foi anteriormente mencionado relativamente à análise da região E1/E2, a análise filogenética das 41 sequências de GBV-C obtidas no decurso deste estudo, referentes à região NS5A/NS5B, teve por base a construção de um alinhamento múltiplo de sequências nucleotídicas, num total de 90, com 884nt de tamanho. A reconstrução de filogenias foi, igualmente, baseada na utilização dos 3 métodos de inferência filogenética referidos em 3.4.1.

Pela observação das três árvores filogenéticas assim reconstruídas verificou-se, genericamente, uma topologia análoga entre as mesmas, com uma organização semelhante da maioria das linhas evolutivas. É, no entanto, de salientar, o facto de que a árvore baseada no algoritmo de NJ apresentou valores de *bootstrap* extremamente baixos ($<75\%$) na pluralidade dos ramos, pelo que grande parte dos grupos monofiléticos identificados, não têm um suporte estatisticamente significativo (Figura 3.22). Pelo contrário, a abordagem Bayesiana possibilitou a obtenção de árvores filogenéticas onde, e porque todas as inferências são feitas tendo em conta as probabilidades *a posteriori*, foi possível aferir que grande parte dos *taxa* foram associados em agrupamentos *major* suportados por valores de $P > 0,9$ (Figura 3.23 A) e, por consequência, consistentes do ponto de vista estatístico.

A segregação de algumas das sequências estudadas, apresentou resultados divergentes daqueles que derivaram da análise da região E1/E2, constatando-se uma maior dispersão das mesmas nas árvores obtidas. Nomeadamente, a sequência 8227NS5B, tida como pertencente a G2^a (por analogia com o estudo efectuado na região E1/E2), não foi incluída em nenhum agrupamento *major* (genótipo), tendo segregado com duas amostras de referência, classificadas como pertencentes aos genótipos 1 (AB003291, com $P=1,00$) e 5 (AY949771), respectivamente. Curiosamente, a sequência AB003291, com a qual a sequência 8227NS5B se encontra filogeneticamente mais próxima, divergiu significativamente das restantes referências de genótipo 1. De igual modo, as sequências 1641NS5B e 0594NS5B, embora relacionadas filogeneticamente ($P=1,00$) com uma amostra de referência de genótipo 1 (AB013500), formaram um agrupamento juntamente com uma sequência de genótipo 6 nas árvores obtidas por análise Bayesiana e pelo critério de Máxima Verosimilhança (Figura 3.23), divergindo das restantes amostras em estudo, classificadas, *a priori*, como pertencentes ao genótipo 1, face aos

resultados obtidos para a região E1/E2. Na árvore de NJ, embora se verifique uma tendência para formação de um agrupamento que incluiu a maioria das sequências de referência do genótipo 1, é também claro o afastamento dessas amostras (1641 e 0954), do grupo das restantes amostras em estudo, classificadas, *a priori*, como pertencentes ao genótipo 1 (sinalizado a cor verde). Para além disso, a sequência utilizada como referência para o genótipo 6 segregou de forma independente quer das amostras em estudo, quer das excedentes sequências de referência.

Em relação à amostra 1631NS5B, quer pela análise efectuada na região E1/E2, quer pelo resultado obtido por NJ, foi congruente a sua associação com o genótipo 2. No entanto, as topologias da árvore Bayesiana e da obtida pelo critério de ML, denotam uma maior proximidade com o grupo designado por G2^a, ainda que deste tenha divergido prematuramente (Figura 3.23).

Por seu lado, a sequência 9317NS5B ocupa uma posição associada, em termos evolutivos, ao “genótipo 1” (pela observação das três árvores filogenéticas obtidas), distinta da anteriormente verificada pela análise da região E1/E2, que resultou na sua inclusão incontestável no genótipo 2. Com um valor de *bootstrap* de 100% (Figura 3.22) e de probabilidade à *posteriori* de 1,00 (Figura 3.23 A), esta sequência sugere uma relação epidemiológica com a designada por 8395NS5B.

Apesar da topologia das três árvores reconstruídas serem quase coincidentes, não ocorreu uma segregação consistente da sequência 6388NS5B com o mesmo agrupamento *major*, dependendo esta do método de inferência filogenética utilizado. Desta feita, observando o filograma da Figura 3.23 B, a referida amostra pareceu manifestar um grau de parentesco evolutivo próximo do agrupamento formado pelo genótipo 3 (embora sem um valor de *bootstrap* que o suporte), ao passo que na análise Bayesiana (com $P=0,92$ no nó de ramificação), à semelhança do observado na árvore de ML, demonstrou contiguidade com uma sequência de referência de genótipo 1.

Por fim, foi possível depreender o agrupamento das restantes amostras em estudo nos principais grupos filogenéticos, G2 (subdividido em G2* e G2^a) e G1 já distinguidos em 3.4.2.



C

G1 - AB013500, AB003291, U36380 (①)

G2 - AB003289, AF031829, AF081782, AF121950, AF309966, AY196904, D87255, D90600, NC_001710, U45966 (②)

G3 - AB003288, AB003290, AB003293, AB008335, AB008342, AF006500, D87262, D87263, D87708, D87709, D87710, D87711, D87712, D87713, D87714, D87715, D90601, U75356, U94695 (③)

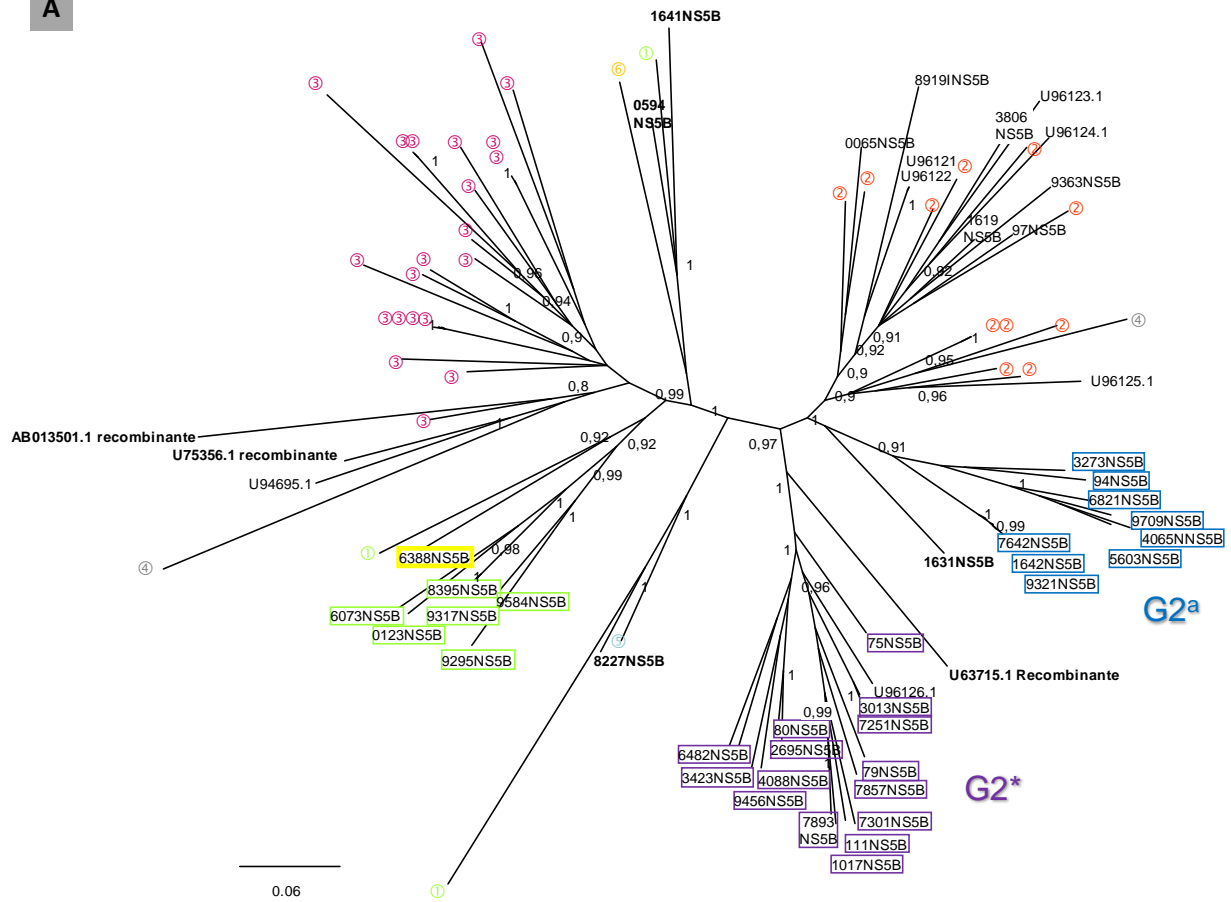
G4 - AB018667, AB021287 (④)

G5 - AY949771 (⑤)

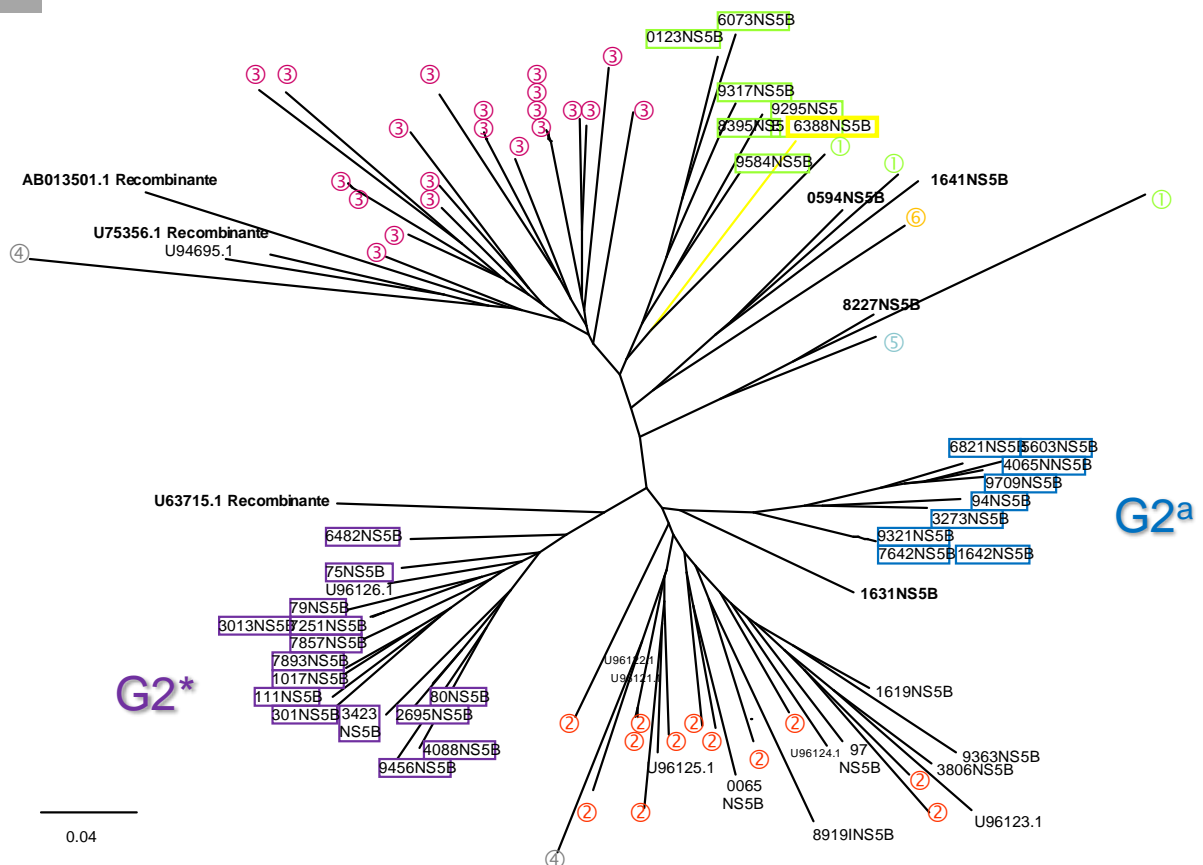
G6 - AB003292 (⑥)

Figura 3.22. Relações filogenéticas entre as 45 estirpes de GBV-C em estudo, baseadas na análise das sequências nucleotídicas da região NS5A/NS5B. As árvores filogenéticas representadas, ambas sem raiz, foram obtidas pelo método NJ, a partir de matrizes de distâncias calculadas pelo método K-2P. Estão também indicados os valores de *bootstrap* (percentagem de 1500 réplicas) considerados significativos acima de 75. A escala representa 1% de distância genética. As sequências das amostras estudadas estão indicadas pela respectiva numeração e com o sufixo NS5B. As sequências de referência parciais relativas à região NS5B têm o prefixo U. **A)** Representação da árvore filogenética obtida. Estão também indicados o número de acesso das respectivas sequências de referência, depositadas nas bases de dados **B)** Correspondência com a árvore filogenética representada em A), onde por questões gráficas e de fácil interpretação, as sequências de referência estão evidenciadas pelo número do genótipo correspondente e com diferentes cores, cuja legenda se encontra em **C)**.

A



B



C

Gen.1 - AB013500, AB003291, U36380 (①)

Gen.2 - AB003289, AF031829, AF081782, AF121950, AF309966, AY196904, D87255, D90600, NC_001710, U45966 (②)

Gen.3 - AB003288, AB003290, AB003293, AB008335, AB008342, AF006500, D87262, D87263, D87708, D87709, D87710, D87711, D87712, D87713, D87714, D87715, D90601, U75356, U94695 (③)

Gen.4 - AB018667, AB021287 (④)

Gen.5 - AY949771 (⑤)

Gen.6 - AB003292 (⑥)

Figura 3.23. Relações filogenéticas entre as 41 estirpes de GBV-C em estudo, baseadas na análise das sequências nucleotídicas da região NS5A/NS5B, 36 sequências de referência de genoma completo relativas aos 6 genótipos reconhecidos de GBV-C, 3 recombinantes e 10 sequências parciais correspondentes à região NS5B (indicadas com o prefixo U). As sequências das amostras estudadas estão alistadas pela respectiva numeração e com o sufixo NS5B. As sequências de genoma completo de referência, por questões gráficas e de fácil interpretação, estão evidenciadas pelo número do genótipo correspondente e com diferentes cores. As árvores filogenéticas radiais representadas, ambas sem raiz, foram obtidas por 2 métodos de reconstrução distintos. **A)** Inferência filogenética baseada numa análise Bayesiana ao longo de 6×10^6 gerações, tendo por base o modelo GTR. Foram considerados significativos os agrupamentos que apresentam valores de probabilidade *a posteriori* $>0,9$. **B)** Árvore filogenética obtida com base no critério de Máxima Verossimilhança. A proporção de sítios invariáveis foi equivalente a 0,45 e o parâmetro $\alpha=0,65$. **C)** Correspondência entre as sequências de referência utilizadas com o respectivo genótipo.

A pesquisa de recombinantes através da análise de tipo *bootscanning* foi aplicada às amostras para as quais a sua associação com agrupamentos monofiléticos nas árvores filogenéticas construídas revelou resultados dúbios, tal como foram casos as sequências 8227NS5B, 1641NS5B, 1631NS5B e 6388NS5B. A análise em questão foi efectuada tendo em conta as sequências correspondentes aos 6 genótipos reconhecidos de GBV-C, e, por outro lado, dado que os genótipos 4, 5 e 6 estão apenas representados por uma única sequência de referência, realizou-se uma segunda análise restringida aos grupos genéticos que evidenciaram maior grau de analogia com as sequências em estudo. Os resultados obtidos com ambas as abordagens foram semelhantes, embora aqueles cuja análise foi mais restrita tenham revelado uma representação gráfica de maior facilidade de interpretação, razão pela qual os primeiros não estão apresentados. Foram considerados significativos os valores de *bootstrap* superiores a 70%.

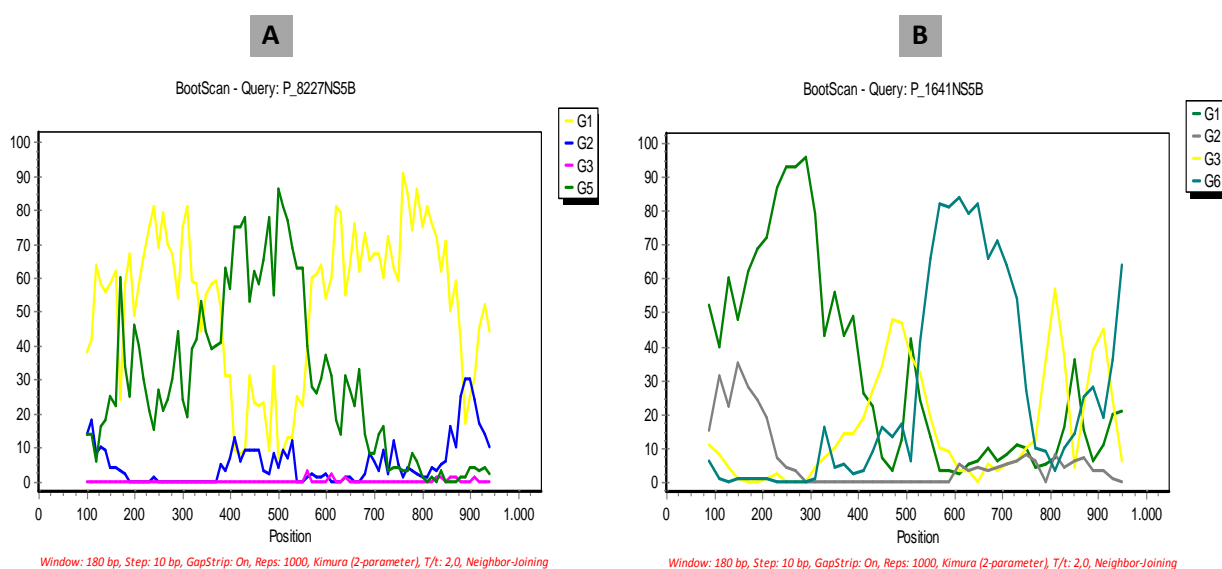


Figura 3.24. Análise de *bootscanning* por *SimPlot* (janela de 180pb, incremento de 10pb e 1000 replicados; inferência filogenética com base no método de *Neighbor-Joining*) da sequência 8227NS5B **A**) e da amostra 1641NS5B **B**). A legenda dos gráficos faz correspondência com os respectivos genótipos em análise. G- corresponde a “genótipo”.

Observando a Figura 3.24 A, verificou-se que a sequência 8227NS5B aparenta uma natureza recombinante com uma estrutura em mosaico do tipo G1/G5/G1, justificando o resultado sugerido pela análise das árvores filogenéticas construídas. A extremidade 3' não apresentou semelhança com nenhuma das sequências de referência utilizadas, pelo que esta região é designada por U (do inglês, *unknown* – desconhecida). Por seu lado, a sequência 1641NS5B (Figura 3.24 B) evidenciou fenómenos de recombinação entre sequências dos genótipos G1 e G6,

apesar de na região central (posição ≈ 400 -500nt) e na região correspondente à extremidade 3', mais uma vez, não ser possível clarificar a sua origem genética.

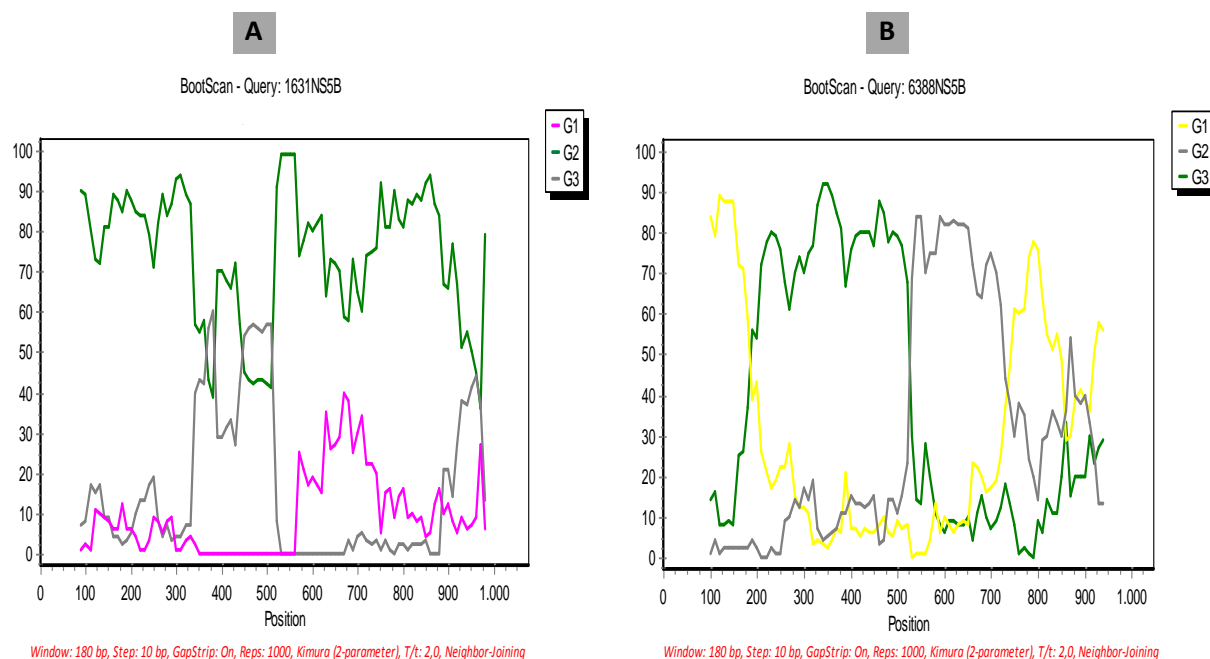


Figura 3.25. Análise de *bootscanning* por *SimPlot* (janela de 180pb, incremento de 10pb e 1000 replicados; inferência filogenética com base no método de *Neighbor-Joining*) da sequência 1631NS5B **A**) e da amostra 6388NS5B **B**). G- corresponde a “genótipo”.

Em relação à sequência 1631NS5B (Figura 3.25 A) verificou-se um grau de semelhança com o genótipo 2, excepto na região entre ≈ 320 -500nt cujo valor de *bootstrap* diminuiu abaixo dos 70% impossibilitando, assim, a identificação da sua origem genética. Finalmente, no que respeita à análise da sequência 6388NS5B, a estrutura em mosaico que sugere a existência de pontos de recombinação é dada por G1/G3/G2/G1, sugerindo uma estrutura recombinante intergenótipos (Figura 3.25 B).

Excluindo as 5 sequências referentes à região NS5A/NS5B para as quais a atribuição de um genótipo foi ambígua (1631E2, 1641E2, 0594E2, 8227E2 e 6388E2), as restantes 36 constituíram grupos monofiléticos. Destas, 30 segregaram com o genótipo 2 (incluindo 9 em G2^a e 15 no “sub-grupo” G2*) (Figuras 3.22 e 3.23). De modo peculiar, as 6 sequências remanescentes analisadas, segregaram apenas com uma sequência de referência de genótipo 1 (Figura 3.26 A).

A formação do sub-grupo G2*, tal como foi definido relativamente à análise da região E1/E2, foi confirmada por uma análise da divergência genética entre as 36 sequências de GBV-C em estudo

através da ferramenta PCOORD (Figura 3.26 B). Por outro lado, o grupo definido como G2^a, evidencia uma proximidade genética significativa com as sequências de genótipo 2, não se denotando como tal, um isolamento expressivo entre os dois conjuntos de sequências, ao contrário do sucedido relativamente a G2*, G1 e G2.

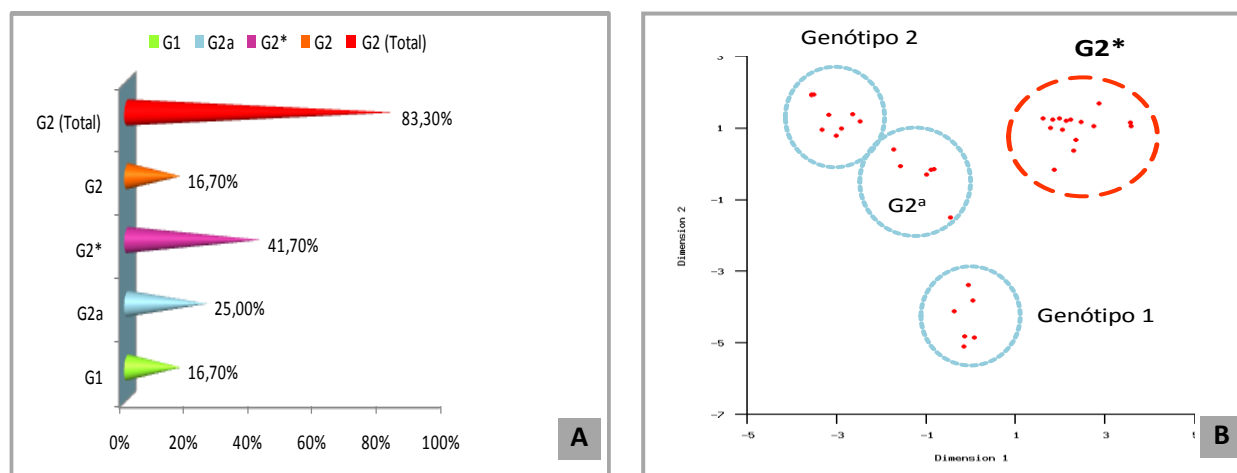


Figura 3.26. A) Representação gráfica da percentagem de estirpes de GBV-C em estudo, amplificadas para a região NS5A/NS5B, agrupadas com os respectivos genótipos. G2 (Total) indica a totalidade de sequências que foram agrupadas neste genótipo (conjunto das sequências em estudo associadas a G2, G2^a e G2*); **B)** Análise obtida pela ferramenta PCOORD para os grupos correspondentes ao genótipo 1, 2 e sub-grupos G2* e G2^a. Os eixos representam as duas dimensões que foram primeiramente extraídas e caracterizam 22,40% das diferenças encontradas entre os 4 grupos.

Com base neste mesmo conjunto de sequências, o cálculo da distância genética obtida quando sequências caracterizadas como possíveis partições do genótipo 2 (considerando os agrupamentos G2, “G2^a” e G2*) foram comparadas entre si, revelou um valor de 10,60% entre G2^a e G2, o menor comparativamente com os restantes grupos envolvidos (Figura 3.27.).

No entanto, uma diferença de 0,40% obtida pelo cálculo da distância genética entre os três grupos, não parece sustentar que existam diferenças significativas entre as sequências em análise, suportando a hipótese de que todas elas pertencem a um mesmo genótipo, em causa, o genótipo 2.

Relativamente aos valores obtidos para a distância genética dentro de cada respectivo grupo em análise (G2, G2* e G2^a), ou seja a distância genética “intragrupos”, constata-se que o grupo G2 é o mais heterogéneo (9,50%), sendo G2^a aquele cujas sequências estão mais próximas geneticamente (com uma distância genética média de 6,90%).

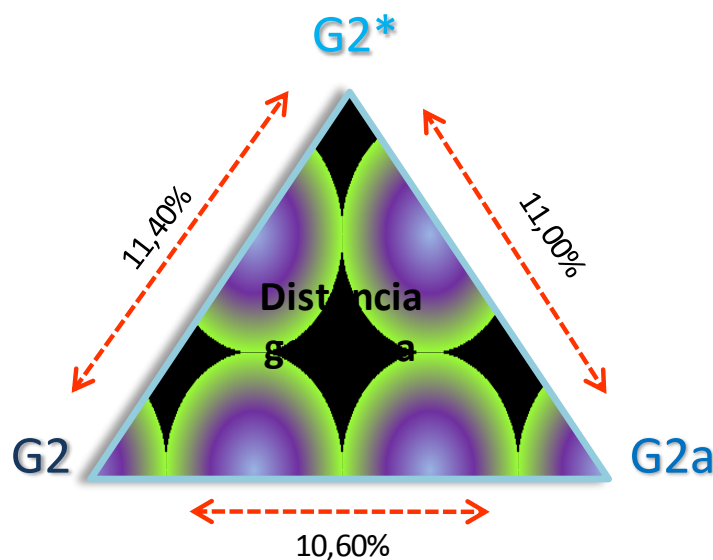


Figura 3.27. Comparação das distâncias filogenéticas, entre os hipotéticos grupos $G2^a$, $G2^*$ e $G2$. As respectivas distâncias encontram-se representadas por setas de diferentes tamanhos.

De igual forma, e por analogia com a análise feita anteriormente pela ferramenta PCOORD, foram calculadas as distâncias genéticas entre os genótipos 1 (considerando o agrupamento hipotético das 6 sequências de GBV-C que segregaram com uma de referência $G1$), genótipo 2 (inclui o conjunto de sequências obtidas para $G2$ e $G2^a$), e ainda o sub-grupo $G2^*$. Verificou-se que os grupos $G2^*$ e $G1$ são aqueles que apresentam maior distância genética entre si (12,40%). A distância média relativa aos 3 grupos em estudo foi de 10,10% (Tabela 3.3 A).

A mesma abordagem foi ainda seguida para a análise das sequências para as quais não foi possível uma caracterização filogenética consistente (ou seja, onde não se verificou uma segregação clara destas sequências, com nenhum dos agrupamentos *major* identificados nas árvores filogenéticas obtidas). Desta forma procurou determinar-se com quais genótipos (individualizando o grupo $G2^*$ das restantes sequências de genótipo 2), estariam mais próximas geneticamente estas sequências.

Na Tabela 3.3 B, encontram-se indicados os resultados da referida análise, denotando-se que a sequência 1641 (e a sequência 0594, por analogia) está a nível genético, mais próxima do grupo relativo ao genótipo 1 ($G1$), tal como a amostra 8227. Por seu turno, as sequências 6388 e 1631 estão geneticamente mais relacionadas com o grupo $G3$ (genótipo 3) e $G1$ respectivamente.

Tabela 3.3. Cálculo das distâncias genéticas entre as amostras indicadas e o respectivo grupo/genótipo em análise. G1 corresponde ao grupo de genótipo 1, G2 ao grupo de genótipo 2 e G3 ao grupo de genótipo 3. Em G1 e G2 foram utilizadas não só sequências de referência como também as de GBV-C em estudo que segregaram com estes grupos. **A)** Tabela relativa às distâncias genéticas entre os genótipos 1, 2, e o grupo G2*a e as sequências indicadas. **B)** Tabela das distâncias genéticas considerando G1, G2, G3, o grupo G2* e as respectivas amostras em análise.

A			B				
Distância genética média (K2P)			Distância genética média(K2P)				
	G2	G2*	Amostras	G1	G2	G2*	G3
G2*	11,20%		1631		5,10%	6,20%	
			1641	1,00%	2,60%		
G1	12,20%	12,4%	8227	4,70%		6,40%	
			6388	5,60%			1,90%

Por analogia à análise efectuada para a região E1/E2 decidiu-se avaliar a razão entre a taxa de substituições nucleotídicas sinónimas (d_S) e não-sinónimas (d_N) para a região NS5A/NS5B. A razão d_S/d_N foi calculada para todas as 36 sequências nucleotídicas de GBV-C amplificadas para esta região (a partir de um alinhamento múltiplo com 882nt), e a média de valor obtido foi de 0,29, o que indica que as sequências estão sobre uma aparente pressão positiva. A mesma razão foi calculada para o grupo G2 em estudo (compreendendo G2a e G2*) e incluindo as sequências de referência agrupadas no genótipo 2 (alinhamento múltiplo com 966nt), de onde resultou um valor médio de $d_S/d_N=0,26$

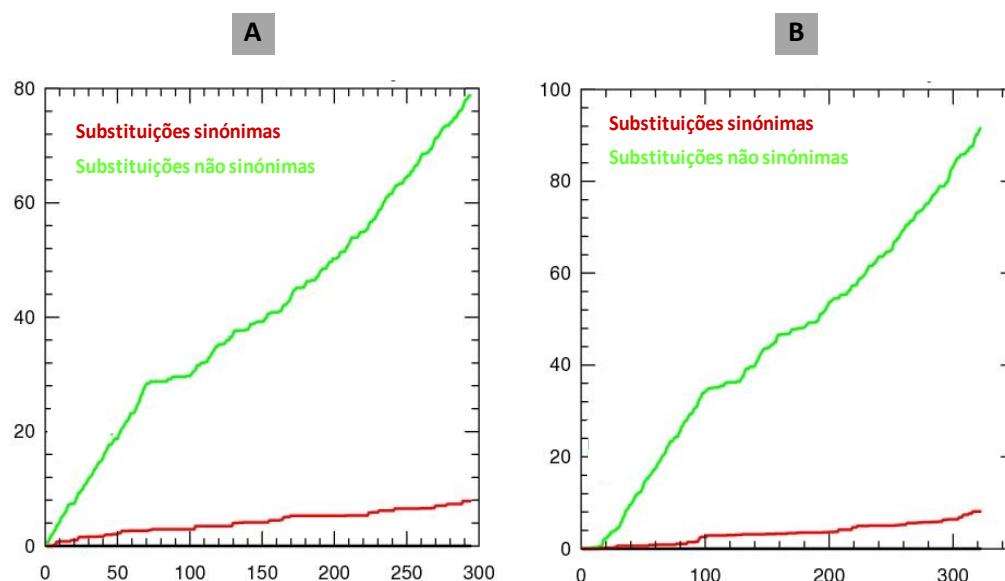


Figura 3.28. Representação gráfica do efeito acumulativo da ocorrência de substituições sinónimas e não sinónimas por codão. Em **A)** estão representados os resultados referentes ao conjunto de 36 sequências de GBV-C analisadas na região NS5A/NS5B, e em **B)**, a representação gráfica obtida para o conjunto de sequências alusivas ao genótipo 2.

Em ambos os casos, verificou-se uma tendência para a ocorrência de substituições não sinónimas indicativo de variabilidade na sequência primária das proteínas (Figura 3.28).

As posições dos aminoácidos foram também avaliadas quanto à sua variabilidade através do cálculo do valor de Entropia de Shannon, tal como anteriormente definido (secção 3.4.2.1). Como grupo alvo (*query*) foram estudadas as sequências pertencentes a G2* (alinhamento com 323aa), uma vez que também no estudo da região NS5A/NS5B do genoma de GBV-C, segregaram de forma particular do genótipo 2.

Como grupo de comparação (*background*) seleccionaram-se as sequências pertencentes ao genótipo 2 (incluindo o conjunto de sequências de GBV-C em estudo e um grupo de sequências de referência G2) num alinhamento múltiplo com 317aa. A conservação da posição de determinados aminoácidos relativamente ao grupo G2* foi evidenciada com valores estatisticamente significativos ($P < 0,005$) relativamente ao resíduo L51 (Figura 3.29).

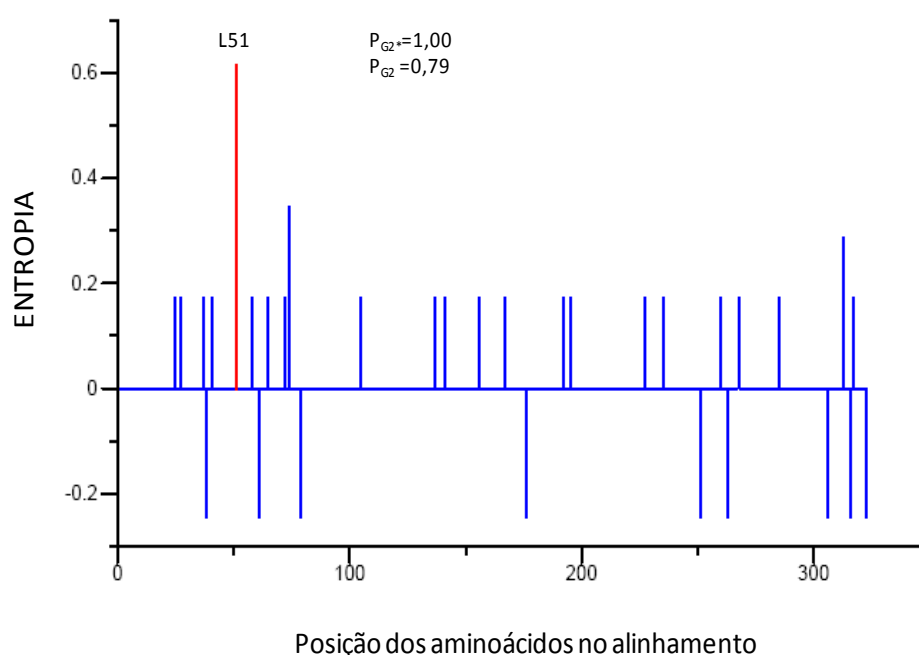


Figura 3.29. Representação gráfica das posições conservadas dos resíduos de aminoácidos, assinaladas por barras vermelhas, com indicação dos valores de probabilidade a estas associados. Análise do conjunto de sequências G2* vs o grupo de sequências alusivas ao genótipo 2.

3.5. Análise semi-quantitativa da virémia de GBV-C por PCR em Tempo Real

3.5.1. Construção de uma recta-padrão

Esta etapa do trabalho realizado teve por base a construção de uma molécula recombinante de DNA plasmídico, derivada de pGEM[®]-T e contendo um inserto de DNA (≈ 250 pb) correspondendo à região 5'UTR, amplificado por RT-PCR a partir de RNA total extraído da amostra 1642 (ver secção 3.3, Tabela 3.2). Este plasmídeo recombinante veio a ser designado pCB1. A confirmação da sua estrutura foi executada como sumariamente descrito na secção 2.8.6. Uma vez purificado, o extracto obtido foi quantificado espectofotometricamente. Face a este resultado ($250\text{ng}/\mu\text{l}$), a utilização da equação $[(\text{quantidade de DNA}(\text{ng}) * 6.022 \times 10^{23}) / (\text{tamanho do fragmento}(\text{pb}) * 1 \times 10^9 * 650)]$, disponível em <http://www.uri.edu/research/gsc/resources/cndna.html>, permitiu a conversão da concentração em ácidos nucleicos deste extracto, expressa em ng de DNA por unidade de volume (μl), no número de moléculas de DNA de cadeia-dupla por unidade de volume ($1,4 \times 10^{11}$ moléculas/ μl).

A natureza semi-quantitativa da análise efectuada resultou do facto de que o procedimento de amplificação entre as amostras biológicas testadas e aquele usado para a construção da recta padrão terem sido semelhantes, mas não idênticos. Apesar das condições de reacção terem sido análogas para ambos os métodos testados (temperaturas de hibridação, *primers* utilizados, quantidade de DNA matriz etc.), a construção da recta padrão foi efectuada a partir de uma molécula de DNA recombinante em cadeia dupla, ao passo que a tentativa de quantificação do número de cópias de genoma de GBV-C nas amostras biológicas foi realizada a partir de cDNA. Com o objectivo de construir uma recta padrão para a semi-quantificação de moléculas de genoma de GBV-C amplificadas por PCR em Tempo Real, o extracto de DNA contendo pCB1 foi diluído de forma seriada variando entre 10^{-1} e 10^{-10} , o que equivaleu a um número de moléculas a alternar entre $1,4 \times 10^{11}$ e 140. Tendo por base o procedimento experimental descrito no capítulo anterior (secções 2.9.2.1 e 2.9.2.2), foram construídas rectas-padrão baseadas nas curvas de amplificação obtidas por PCR em Tempo Real recorrendo a dois métodos distintos de detecção dos produtos de amplificação: utilização de uma sonda *TaqMan* e do agente intercalante *SYBR Green I* (Figura 3.30 A). A sonda *TaqMan*, bem como o par de *primers* utilizado (Tabela 2.11, Capítulo 2) foram descritos num estudo anterior (Hattori *et al.*, 2003).

Os ensaios realizados com a referida sonda conduziram à amplificação de um fragmento da 5'UTR com $\approx 130\text{pb}$ (tal como confirmado após separação dos produtos de reacção em gel de agarose, resultados não apresentados) a partir do extracto de DNA contendo pCB1 (diluído em série).

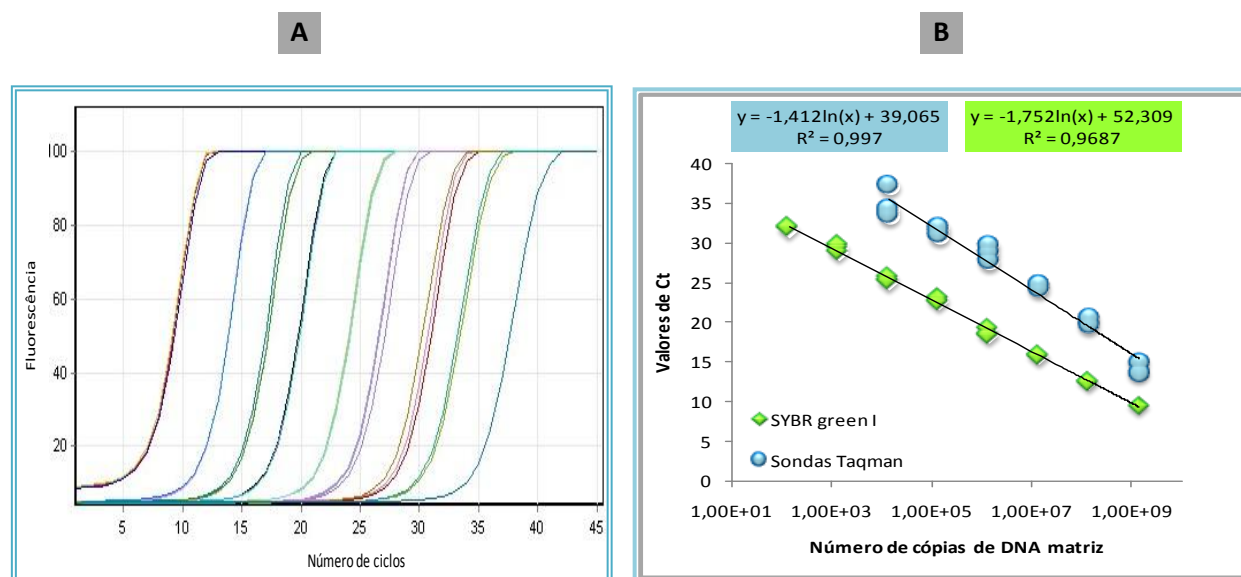


Figura 3.30. A) Amplificação por PCR em Tempo Real de um fragmento de DNA (da 5'UTR) usando como matiz as diluições de seriadas do plasmídeo pCB1, por recurso ao corante *SYBR Green I* na detecção do DNA amplificado. Foram efectuadas três réplicas de cada diluição (curvas de amplificação sobrepostas). **B)** Representação gráfica das rectas de amplificação obtidas a partir dos valores de Ct, que resultaram dos ensaios com uma sonda *TaqMan* (azul) – e dos ensaios com o corante *SYBR Green I*, indicados com marcadores de cor verde.

A definição dos valores de Ct (ou *Cycle Threshold* – corresponde ao ponto que detecta o ciclo no qual a reacção atinge o limiar da fase exponencial) obtidos para cada uma das diluições de pCB1 analisadas, permitiu a construção de um gráfico em que, independentemente do facto da detecção de DNA amplificado ter resultado do sinal de fluorescência emitido em consequência da intercalação de *SYBR Green I*, ou da hidrólise da sonda *TaqMan* durante a polimerização das moléculas de DNA (Figura 3.30 B), este é expresso em função da quantidade de pCB1. É, no entanto, de salientar, que os resultados obtidos por ambos os métodos não foram exactamente coincidentes. Tal facto poderá ter resultado de diferenças a nível da sensibilidade da detecção de DNA amplificado tendo esta, nas condições experimentais utilizadas, sido maior aquando da utilização de *SYBR Green I* (detecção da amostra diluída até 10^{-11}). Inversamente, constatou-se uma perda de sensibilidade (detecção da amostra diluída até 10^{-9}) nas situações com utilização da sonda *TaqMan*.

3.5.2. Quantificação relativa do genoma de GBV-C nas amostras biológicas estudadas

Nove dos extractos de cDNA preparados a partir de RNA extraídos de amostras biológicas seleccionadas aleatoriamente (com designação de A-I, correspondentes a 9584, 6073, 111, 75, 79, 9709, 7642, 7857 e 3068, respectivamente), foram submetidas às mesmas condições de amplificação descritas para a construção das rectas padrão. Nos ensaios realizados com *SYBR Green I*, as amostras estudadas foram diluídas com um factor de diluição 1:50 de forma a diminuir a presença de moléculas de elevada massa molecular, observadas por análise dos produtos de amplificação em gel de agarose (não apresentado), uma vez que estas moléculas interferem no processo de detecção por *SYBR Green I* já que este se liga a qualquer DNA em cadeia dupla. Por seu lado, aquando da utilização de sonda específica a diluição utilizada foi de 1:10.

A partir da regressões lineares das rectas padrão, e com os valores de Cts auferidos desta análise, foi possível calcular o número de cópias de DNA por reacção, os quais estão indicados na Tabela 3.4.

Tabela 3.4. Análise semi-quantitativa do número de cópias de DNA por reacção estimados a partir dos declives das respectivas rectas padrão.

Amostras	Ct (<i>SYBR Green I</i>)	Nº cópias de DNA por reacção	Ct (Sonda <i>TaqMan</i>)	Nº cópias de DNA por reacção
A	28,5	1,19E+04	-	-
B	26,64	4,40E+04	-	-
C	-	-	-	-
D	27,93	1,77E+04	31,28	1,63E+06
E	27,09	3,20E+04	29,86	3,67E+06
F	27,32	2,73E+04	30,06	3,27E+06
G	28,98	8,46E+03	32,82	6,78E+05
H	23,69	3,52E+05	27,32	1,56E+07
I	39,33	5,75E+00	37,06	6,03E+01

Nas reacções de PCR em Tempo Real relativas às amostras em estudo, não foi detectado sinal de amplificação para as amostras A, B (pelo método com sonda *TaqMan*) e C (em ambos os métodos), apesar de terem sido testadas alternativas nas condições de amplificação, como diminuição da temperatura de hibridação dos *primers* de 60° para 55°, e, aumento da quantidade de sonda utilizada por reacção, de 1µl para 2µl. Com base nestes resultados, colocou-se a hipótese de que a sonda pudesse não hibridar convenientemente com a sequência dos amplicões

obtidos (cuja presença foi verificada em gel de agarose), impedindo assim a sua degradação pela enzima *Taq* polimerase e justificando a ausência de sinal. Desta forma, foram realizadas novas reacções de *heminested-PCR*, como descrito na secção 2.9.2.3 (Capítulo 2), para amplificação do fragmento que inclui parte da 5'UTR (≈ 250 nt), a partir do cDNA das amostras A, B e H (esta última foi escolhida por ter sido aquela onde foi detectado o número mais elevado de cópias de DNA por ambos os métodos de detecção), de modo a poder comparar as sequências dos amplicões em causa, com a sequência nucleotídica da sonda verificando, ou não, a identidade entre ambas.

As condições de amplificação foram comuns a ambas as etapas de PCR e corresponderam às descritas para a amplificação da região 5'UTR indicadas na Tabela 2.3 (Capítulo 2). O produto obtido para as três amostras foi submetido a sequenciação e as sequências daí resultantes comparadas com a da sonda *TaqMan* utilizada. Com base na referida análise, foi possível identificar a presença de 5 posições nucleotídicas distintas nas amostras A e B, evidenciando desta forma, que as sequências obtidas e a da sonda *TaqMan* não foram coincidentes (*mismatch*) (Figura 3.31). A amostra H não revelou a presença de qualquer posição nucleotídica discordante com a sequência respectiva à sonda (CCCGACGTCAGGCTCGTCGTAA).

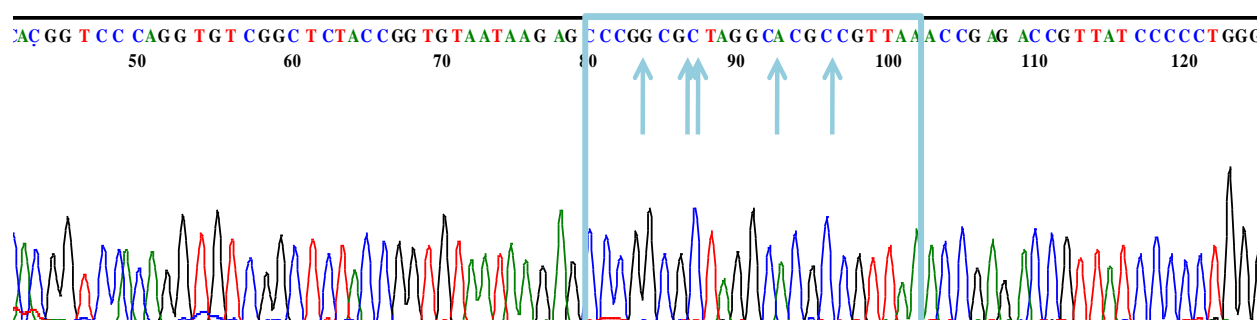


Figura 3.31. Cromatograma respectivo à sequenciação da amostra com indicação das bases que revelaram posições nucleotídicas não homólogas com a sequência nucleotídica da sonda *TaqMan* (setas azuis).

Por seu turno a amostra C foi tentativamente amplificada por *nested-PCR* (como descrito em 2.5.1, pelo sistema *puReTaq Ready-To-Go PCR Beads*) sem sucesso, razão pela qual não se procedeu à sua sequenciação.

4. DISCUSSÃO DE RESULTADOS

PREVALÊNCIA DE GBV-C REGIÃO METROPOLITANA DE LISBOA: ESTUDO DE UM GRUPO DE INDIVÍDUOS COM ELEVADO RISCO DE INFECÇÃO POR VÍRUS DE TRANSMISSÃO PARENTÉRICA

A descoberta do GBV-C e a sua associação inicial enquanto agente etiológico de hepatites nos humanos, conduziu à realização de inúmeros estudos que pretenderam esclarecer diversos aspectos da biologia deste vírus. São destes exemplo, a forma como podem influenciar o estado de saúde de diferentes grupos populacionais, tais como os UDIs e os indivíduos hemodialisados e transplantados (essencialmente renais e hepáticos), onde a sua prevalência é elevada (Reshetnyak *et al.*, 2008). Os dados epidemiológicos até hoje obtidos, apontam a via parentérica como a principal via de transmissão deste vírus (Feucht *et al.*, 1997; Lefrère^a *et al.*, 1999; Björkman *et al.*, 2001; Stapleton 2003). No mesmo encadeamento, e apesar do facto da transmissão do GBV-C por via percutânea ser eficaz, as vias sexual e materno-fetal (transmissão vertical) são relativamente frequentes e com significado proeminente em termos epidemiológicos (Ibáñez *et al.*, 1998; Nerurkar *et al.*, 1998, Scallan *et al.*, 1998; Ohto *et al.*, 2000; Mathet *et al.*, 2003; Handelsman *et al.*, 2007). Por outro lado, destaca-se ainda a elevada tendência para a ocorrência de co-infecções com outros vírus transmitidos pelas mesmas vias, distinguindo-se, de entre outros, o HIV e o HCV (Rendina *et al.*, 2001; Qurishi *et al.*, 2003; Berzsenyi^a *et al.*, 2005; Mehrishi *et al.*, 2005; Barusruk *et al.*, 2006; Mosam *et al.*, 2007, Berzsenyi *et al.*, 2009).

Neste contexto, e tendo em conta o facto de que a informação disponível sobre a presença de GBV-C na população Portuguesa ser claramente escassa, o presente trabalho propôs uma análise, não só de carácter epidemiológico, como também de caracterização genética das estirpes de GBV-C encontradas em Portugal. O trabalho aqui apresentado centrou-se no estudo de uma população caracterizada por elevadas taxas de infecção por outros vírus de transmissão parentérica (HIV e/ou HCV), e circunscrevendo residentes da área metropolitana de Lisboa. A amostragem em análise, envolveu um grupo de indivíduos de nacionalidade portuguesa (maioria) e ainda um número considerável (cerca de 28,50%) de indivíduos de etnia africana. De um total de 214 indivíduos é de salientar o facto de 151 destes, reportarem o consumo de drogas por via endovenosa como a principal via de transmissão de HIV/HCV.

A estimativa do contacto de uma população com o GBV-C pode ser efectuada através da detecção da presença do RNA viral no plasma/soro (Erker^b *et al.*, 1998; Odeh *et al.*, 2005; Watanabe *et al.*, 2008), a qual indica a presença do vírus na circulação sanguínea periférica (virémia), e funciona, portanto como um marcador de infecção activa. Em alternativa, podemos

recorrer à detecção da resposta imune contra este vírus através da realização de testes serológicos. Estes permitem a detecção de anticorpos dirigidos contra a proteína E2 do invólucro viral (anti-GBV-C), funcionando estes anticorpos, como marcadores de infecção passada (Lefrère^a *et al.*, 1999; Giulivi *et al.*, 2000; Reshetnyak *et al.*, 2008). Os dois marcadores (RNA/anticorpos) não são, geralmente, sobreponíveis, sugerindo que os anticorpos anti-GBV-C sejam neutralizantes (Rey *et al.*, 2000; George e Varmaz, 2005; Reshetnyak *et al.*, 2008). A virémia de GBV-C é determinada pela detecção do RNA viral por RT-PCR (*nested* ou *heminested*-PCR) (Dawson *et al.*, 1996; Erker^b *et al.*, 1998; Souza *et al.*, 2006; Reshetnyak *et al.*, 2008), através da qual são amplificadas sequências conservadas do genoma viral (Souza *et al.*, 2006). Os primeiros estudos que envolveram a detecção de GBV-C utilizavam *primers* delineados para a amplificação de fragmentos internos às regiões codificantes das proteínas não-estruturais NS3 e NS5A (Muerhoff *et al.*, 1997; Simons *et al.*, 1995; Ibáñez *et al.*, 1998). No entanto, uma vez que a sensibilidade da detecção desejada depende, em grande parte, da hibridação dos *primers* usados com a matriz de ácidos nucleicos que se pretende amplificar, actualmente recorre-se, em alternativa, à utilização de *primers* complementares à região mais conservada do genoma viral, a 5'UTR (Xiang *et al.*, 2005; Muerhoff *et al.*, 2006; Souza *et al.*, 2006; Baggio-Zappia e Granato 2009). Respeitando este procedimento, foi detectada virémia de GBV-C em 87 de 214 amostras analisadas, revelada através da amplificação, por *nested*-PCR, da 5'UTR sob a forma de fragmentos de DNA com aproximadamente 250pb. Tal como se poderia antecipar, foi registada uma elevada prevalência de infecções activas por GBV-C na população em estudo (40,65%). Este valor é aproximado ao reportado em diferentes partes do globo (Yan *et al.*, 2000; Rey *et al.*, 2000; Berzsenyi^a *et al.*, 2005; Hekmat *et al.*, 2008). De igual forma, a taxa de infecção obtida no círculo de indivíduos consumidores de drogas injectáveis, considerado como um dos principais grupos expostos à infecção pelo GBV-C, foi surpreendente pelo alto valor registado (78,16%), por comparação com os dados epidemiológicos, revistos por Baggio-Zappia e Granato, que indicam valores entre os 3% e 58% (Baggio-Zappia e Granato, 2009). Em Portugal, no final da década de 1990, tinha sido já evidenciada uma percentagem de infecção em indivíduos toxicodependentes de 36% (Carvalho *et al.*, 1999) pelo que, os valores obtidos no presente estudo, sugerem um aumento substancial da prevalência de GBV-C neste grupo de risco.

O potencial impacto clínico das infecções por GBV-C tem sido alvo de exaustivas investigações. Contrariando as perspectivas iniciais que apontavam o GBV-C como um agente de hepatites

virais “não A-E” (Tameda *et al.*, 1996), os dados mais recentes não o associam a nenhuma doença em particular (Reshetnyak *et al.*, 2008). No entanto, o potencial efeito de modulação negativa que o GBV-C parece exercer na replicação e progressão das infecções pelo HIV, tem sido alvo de alguma controvérsia no meio científico (Li^a *et al.*, 2006; Sheng *et al.*, 2007; Nunnari *et al.*, 2003; Handelsman *et al.*, 2007; Mosam *et al.*, 2007). Actualmente, pouco ainda se sabe sobre o(s) possível(eis) mecanismo(s) através do qual(uais) a infecção por GBV-C potencia um efeito benéfico sobre os indivíduos co-infectados com o HIV (George e Varmaz, 2005; Baggio-Zappia e Granato, 2009). A co-infecção com GBV-C aparenta exercer um efeito protector nos indivíduos seropositivos para HIV, diminuindo a velocidade de progressão para SIDA e, consequentemente, a taxa de mortalidade associada à infecção pelo HIV (Stapleton *et al.*, 2004; Berzsenyi^a *et al.*, 2005; Baggio-Zappia e Granato, 2009). A melhoria de sobrevida parece estar confinada àqueles que têm uma elevada virémia de GBV-C, facto que tem sido demonstrado em estudos aplicados a uma proporção significativa de indivíduos infectados (Tillmann *et al.*, 2001; Xiang *et al.*, 2001; Williams *et al.*, 2004). Uma redução na severidade da doença hepática provocada pelo HCV tem também sido, também, sugerida por alguns autores que relacionaram as co-infecções pelos vírus GBV-C/HCV ou GBV-C/HCV/HIV com o grau de inflamação a nível do fígado (Strauss *et al.*, 2002; Berzsenyi *et al.*, 2007; Barbosa *et al.*, 2009).

O GBV-C tem uma distribuição global nos indivíduos infectados com HIV, a qual varia entre 20-62% (Rey *et al.*, 2000; Rendina *et al.*, 2001; Baggio-Zappia e Granato 2009). No presente estudo a prevalência da infecção por GBV-C foi relativamente elevada nos indivíduos co-infectados com HIV e/ou HCV, comparativamente à taxa de mono-infecção pelo GBV-C, estimada em 22,99%. A presença de co-infecções foi detectada em 19,54% de indivíduos seropositivos para HIV e 16,09% dos seropositivos para HCV, enquanto que 41,38% das infecções revelaram a presença dos 3 vírus (GBV-C/HIV/HCV). Estes valores são compatíveis com o facto de estes vírus partilharem vias de transmissão idênticas (Rendina *et al.*, 2001; Amarante *et al.*, 2004; Hoseini *et al.*, 2007; Ramezani *et al.*, 2008).

Estudos realizados no âmbito da transmissão do GBV-C por via sexual, entre indivíduos co-infectados com HIV, demonstraram já uma clara propagação da infecção de GBV-C por esta via, registando valores próximos de 47% (Berzsenyi^b *et al.*, 2005). Assim, a percentagem de indivíduos portadores de HIV e simultaneamente infectados com GBV-C, provavelmente, por contacto sexual (registada no decurso deste trabalho com o valor de 32,7%) é comparável a outros previamente registados de 16,5% (Ibáñez *et al.*, 1998) e 24% (Sawayama *et al.*, 1999)

sublinhando a importância da transmissão viral por esta via (Scallan *et al.*, 1998; Frey *et al.*, 2002). Os resultados obtidos podem, no entanto, expressar de certa forma um efeito dúbio, pelo facto de diferentes vias de transmissão poderem contribuir, de forma combinada, para a dispersão deste vírus nos humanos.

Clinicamente já foi reconhecida uma relação inversa entre a carga viral do GBV-C e a carga viral de HIV nos indivíduos co-infectados, sendo a primeira considerada como um dos putativos factores que sugerem o efeito benéfico, que advém da co-infecção, na progressão para a SIDA (Shankar *et al.*, 2008; Li^a *et al.*, 2006; Maidana-Giret *et al.*, 2009). Tendo sido testada a associação com outras variáveis, nomeadamente, sexo, etnia, carga viral de HIV e contagem de células CD4⁺ (tendo em conta os dados epidemiológicos disponíveis), não se verificou uma associação significativa entre as mesmas e a virémia de GBV-C, análise também evidenciada por Hekmat (Hekmat *et al.*, 2008). No que respeita à não constatação de uma associação significativa entre a presença de GBV-C e a carga viral de HIV, estatisticamente os resultados auferidos podem reflectir o reduzido tamanho da amostra e/ou a variação no estágio de infecção com HIV. No entanto, no presente trabalho ficou demonstrada uma associação significativa ($P < 0,05$) entre a virémia de GBV-C e a idade dos indivíduos infectados, nomeadamente na faixa etária ≥ 25 , resultado que está em concordância com os dados reportados por Rey e Mphahlele, onde foi mostrado que GBV-C infecta população adulta jovem (Rey *et al.*, 2000; Mphahlele *et al.*, 1999). Tal como já anteriormente evidenciado (Christensen *et al.*, 2003; Asim *et al.*, 2008), o presente trabalho revelou, ainda, uma relação estatisticamente significativa entre a presença de virémia de GBV-C com o grupo de utilizadores de drogas injectáveis.

CARACTERIZAÇÃO GENÉTICA DAS ESTIRPES VIRAIS DE GBV-C

O GBV-C existe como uma colecção de variantes estreitamente relacionadas do ponto de vista genético, as quais se associam em agrupamentos. Estes são designados genótipos e incluem estirpes virais que diferem em aproximadamente 11-14% da sua sequência nucleotídica (Smith, 1997; Simmonds^b, 2001; Muerhoff *et al.*, 2005; Berzsenyi^a *et al.*, 2005). A infecção por GBV-C é comum nos humanos, distribuindo-se por todo o globo tendo sido já reportados casos de detecção do vírus em populações indígenas na Papua Nova Guiné (Mison *et al.*, 2000), África Central (Tanaka *et al.*, 1998), e América do Sul (Tanaka *et al.*, 1998). Até à data, os 6 genótipos reconhecidos de GBV-C são caracterizados pela forte associação a localizações geográficas específicas (Reshetnyak *et al.*, 2008). Esta distribuição geográfica parece estar relacionada com a

co-evolução dos vírus com os humanos durante as migrações ao longo da História, sugerindo que o GBV-C seja um vírus ancestral (Smith *et al.*, 1997; Pavesi, 2001; Naito e Abe, 2001). De facto, análises filogenéticas de diferentes isolados de GBV-C, identificadas a nível global, demonstraram que a diversidade entre os isolados africanos deste vírus é maior do que a existente dentro dos outros grupos geográficos, sugerindo como tal uma origem africana, e, coincidindo esta, com o período das migrações humanas para fora de África há mais de 100.000 anos (Tanaka *et al.*, 1998; Pavesi, 2001; Simmonds^a, 2001; Simmonds^b, 2001). Para além disso, a distância genética entre variantes de GBV-CPZ encontrada nos chimpanzés (das subespécies *Pan troglodytes troglodyte* e *Pan troglodytes verus*) e qualquer das estirpes detectadas nos humanos é maior do que as distâncias genéticas quando estirpes de diferentes genótipos de GBV-C (humanos) são comparadas entre si. Tal facto reforça, claramente, a teoria da co-evolução (Birkenmeyer *et al.*, 1998; Charrel *et al.*, 1999; Simmonds^b, 2001).

Consistente com esta possibilidade, os genótipos 1 e 5 são encontrados em África, ao passo que os genótipos 2, 3 e 4 reflectem as três principais rotas da migração humana a partir do continente africano para a Europa (Tanaka *et al.*, 1998; Pavesi, 2001; Relethford *et al.*, 2008; Stanyon *et al.*, 2009; Cann, 2001), onde prevalece o genótipo 2, e norte e sul asiático onde se podem encontrar os genótipos 3 (Lu *et al.*, 2000) e 4 (Naito *et al.*, 2000), respectivamente. O genótipo 2 inclui também isolados da América do Norte, América do Sul, Japão, Paquistão e África Ocidental (revisto por Baggio-Zappia e Granato, 2009). O genótipo 3 é igualmente encontrado na América do Sul, consistente com a provável rota tomada pelos primeiros habitantes desta área, que aí chegaram depois de atravessarem a massa de terra hoje em dia submersa sob o estreito de Bering. Embora os genótipos 1-4 possam ser encontrados no Japão, a maioria dos isolados corresponde ao genótipo 3 (Naito e Abe, 2001). A descrição de um putativo genótipo 6 em estirpes virais identificadas em indivíduos Indonésios foi recentemente evidenciada por Muerhoff e colaboradores (Muerhoff *et al.*, 2006).

A existência de múltiplos genótipos de GBV-C levou a que vários autores sugerissem que estas pudessem diferir do ponto de vista biológico e, por exemplo, poder afectar de forma diferencial a progressão para a doença nos indivíduos seropositivos para HIV co-infectados com GBV-C. De facto, num estudo recente foram observadas importantes diferenças relacionadas com os genótipos de GBV-C em doentes co-infectados com os vírus HIV/HCV, onde a infecção com o genótipo 2 de GBV-C foi associada com maior contagem de células CD4⁺ comparadas com a infecção com isolados de genótipo 1 do mesmo vírus (Schwarze-Zander *et al.*, 2006). Conclusões semelhantes foram registadas por Muerhoff e colaboradores, ao notarem que a

contagem de células CD4⁺ foi mais baixa em indivíduos infectados com o subtipo 2a do que aqueles infectados com o subtipo 2b (Muerhoff *et al.*, 2003), correspondendo estes a subgrupos específicos de estirpes virais do genótipo 2. Embora outros factores não medidos/quantificados (por exemplo, diferenças a nível da população estudada, factores relacionados com o hospedeiro, introdução de terapia anti-retroviral, distintos níveis de linfotropismo dos diferentes genótipos de GBV-C presentes e/ou os subtipos de HIV associados) possam ter influenciado tais resultados, estas observações poderão ter importantes implicações para a compreensão das interacções moleculares entre o GBV-C e HIV, e as consequências de uma infecção com GBV-C na gestão da co-infecção de doentes com HIV/HCV submetidos a terapia anti-retroviral e/ou com interferão. A extensão da diversidade global do GBV-C é, ainda assim, limitada, dado que existe ainda reduzida informação disponível para um grande número de áreas do globo.

Na maioria dos estudos descritos na literatura, a identificação de novos genótipos recorreu- à análise filogenética da 5'UTR, uma vez que o estudo de pequenos fragmentos dispersos ao longo da maior parte da região codificante resulta, normalmente, em filogenias incongruentes (Muerhoff *et al.*, 1996; Muerhoff *et al.*, 1997; Smith *et al.*, 1997; Naito *et al.*, 1999; Sathar *et al.*, 1999; Muerhoff *et al.*, 2006). No entanto, embora estes trabalhos tenham mostrado que a análise da 5'UTR permite uma segregação das sequências das estirpes GBV-C sob a forma de genótipos independentes, os amplicões obtidos para esta região têm um tamanho limitado (aproximadamente 250pb), o que restringe a quantidade de informação genética disponível para a reconstrução de filogenias.

Enquanto os genótipos de HCV podem ser distinguidos por análise filogenética de uma variedade de regiões genómicas de tamanhos tão pequenos quanto 222nt, tal não se verifica para o GBV-C. Estudos baseados em pequenos fragmentos de regiões codificantes, tais como as que codificam as proteínas NS3 (Simons *et al.*, 1995; Masuko *et al.*, 1996; Heringlake *et al.*, 1996; Muerhoff *et al.*, 1997; Ibáñez *et al.*, 1998), ou NS5A e NS5B (Muerhoff *et al.*, 1997; Viazov *et al.*, 1997) falharam na produção de árvores filogenéticas congruentes, dado que a análise da maioria destas regiões não sustentam o mesmo comportamento/relação filogenética apresentados pelos estudos com sequências do genoma completo do vírus (Smith *et al.*, 2000). Por essa razão, a genotipagem das variantes do GBV-C, quando a totalidade da sequência viral não está disponível, é feita com base na análise parcelar da 5'UTR e/ou da região codificante da proteína E2, sendo estas as secções do genoma que permitem genotipagens coerentes com as efectuadas com base na totalidade da sequência deste vírus (Smith *et al.*, 2000; Muerhoff *et al.*, 2005).

Neste trabalho foi proposta, *a priori*, a caracterização genética das estirpes circulantes de GBV-C na área metropolitana de Lisboa, com base na análise filogenética de três regiões subgenómicas: 5'UTR/E1, E1/E2 e NS5A/NS5B. Numa situação ideal, este estudo deveria ter envolvido o estudo de sequências genómicas completas. No entanto, tal não seria exequível devido às limitações impostas pelo tempo útil para a realização da componente prática desta dissertação, e pelos elevados custos associados. A análise simultânea destas três regiões foi justificada não só pela quantidade de informação genética resultante, como pela possibilidade de se verificar até que ponto os resultados de genotipagem poderiam ser influenciados pela análise filogenética das três regiões em separado. Por outro lado, questionou-se até que ponto as regiões genómicas seleccionadas poderiam estar sujeitas a pressões selectivas distintas. Esta análise envolveu, sensivelmente, metade das amostras onde foi detectada virémia de GBV-C.

Uma das regiões estudadas, designada por 5'UTR/E1, foi amplificada sob a forma de um fragmento com aproximadamente 1,2kpb que inclui parte da 5'UTR (≈ 230 nt) e ≈ 940 nt da região que codifica a glicoproteína E1 da superfície viral (a qual compreende, na sua totalidade, aproximadamente 1080nt). A amplificação, por *heminested*-PCR, desta região alvo foi conseguida em 27 das 30 amostras testadas, após optimização do protocolo de amplificação. A amplificação de 5'UTR/E1 não foi conseguida a partir de três das amostras analisadas (identificadas como 9295, 3806 e 7888). Uma quantidade limitante de DNA matriz (em indivíduos com baixa carga viral) ou degradação parcial do RNA viral, resultante da exposição das amostras em estudo a repetidos ciclos de congelação/descongelação, poderá ter condicionado o sucesso conseguido. Adicionalmente, a não amplificação de sequências virais a partir das amostras em questão poderá ainda ter dependido da existência de polimorfismos no genoma viral, os quais condicionam a hibridação entre os *primers* utilizados e a região alvo a amplificar. A amplificação de sequências virais a partir de cDNA preparado a partir de RNA viral extraído das amostras em questão, foram repetidamente testadas com novas condições de amplificação, embora sem êxito, traduzida pela consistente ausência de amplificação ou fraco rendimento na quantidade de produto obtido.

Seis dos 27 produtos obtidos foram sequenciados. Em cinco destes amplicões (referentes às amostras 1642, 0594, 3273, 9709 e 6821), as reacções de sequenciação efectuadas com o *primer* UTRForI terminaram precocemente, tendo resultado sequências com tamanhos a variar entre os 140nt (a de menor tamanho) e um máximo de 335nt. A excepção ocorreu apenas para a

sequência identificada por 4088, cujo resultado obtido com ambos os *primers* utilizados (UTRForI e R1rev3) possibilitou a construção de uma sequência de consenso com 1030nt. Para todas as demais sequências analisadas, as dificuldades técnicas que impossibilitaram a obtenção de bons resultados de sequenciação, mantiveram-se mesmo após a substituição do *primer* UTRForI por outro, designado E1rev4, o qual foi delineado com o intuito de melhorar os resultados previamente obtidos. Consequentemente, devido ao pequeno tamanho das sequências obtidas, a caracterização do genoma de GBV-C baseada na análise desta região do genoma viral foi abandonada, não tendo prosseguido a sequenciação das restantes 21 amostras previamente amplificadas. Esta opção foi tomada tendo em consideração o factor tempo, os objectivos propostos neste trabalho, e os esforços que teriam de ser implementados de forma a permitir ultrapassar os problemas técnicos encontrados.

Os genomas de muitos vírus de RNA, incluindo o GBV-C, contêm regiões que formam estruturas internas por emparelhamento de bases complementares, dando origem a estruturas secundárias com formas de gancho-de-cabelo (*stem-loop*), as quais participam nos processos de regulação da replicação e tradução, este último devido à ligação dos ribossomas a um IRES (*Internal Ribosome Entry Site*) (Simmonds e Smith 1999).

Uma das hipóteses apontadas como justificação para a paragem da polimerase de DNA durante as reacções de sequenciação foi o facto de, aquando o processo de amplificação cíclica, as moléculas de DNA em cadeia simples pudessem assumir estruturas secundárias e terciárias impossibilitando, assim, a progressão da polimerização efectuada por esta enzima. Como tal, decidiu-se investigar a forma como a ocorrência das mesmas poderia ter condicionado os resultados obtidos, tendo por base os modelos que parecem descrever o conjunto de estruturas secundárias incluídas na região em causa (Simons *et al.*, 1996; Thurner *et al.*, 2004). Os métodos convencionais que prevêem a formação de estruturas secundárias em moléculas de RNA são computacionalmente intensivos e são, a maioria das vezes, baseadas no cálculo das energias livres mínimas (Energia de Gibbs ou ΔG°) para cada possível conFiguração de uma dada sequência de RNA (Simmonds e Smith, 1999). A avaliação da estabilidade termodinâmica das estruturas resultantes, revelou valores de ΔG° negativos indicativos de formação de estruturas secundárias potencialmente estáveis.

Com base nas estruturas secundárias formadas (Figura 3.8), designadas por HG6 e HG7, tal como descrito por Thurner e colaboradores (Thurner *et al.*, 2004), foi possível identificar com relativa exactidão os pontos de paragem das sequências obtidas relativamente às amostras 0594

na região HG6, enquanto que para 6821, 1642 e 9709 estas localizaram-se na estrutura correspondente a HG7. No que diz respeito à sequência 3723 (5'UTR/E1), o extremo 3' da mesma foi localizado entre as regiões assinaladas como HG6 e HG7. Esta região, que inclui um possível IRES, é altamente estruturada e caracterizada pela sua hibridação parcial com sequências a montante de HG3 (ver Figura 3.7). Por esta razão, esta região de estrutura complexa, deverá ser suficientemente estável para justificar a interrupção abrupta do processo de sequenciação. Considerando que nas sequências analisadas a formação de estruturas secundárias deverá ser análoga à descrita graficamente na Figura 3.8, ficou demonstrado que as sequências respectivas às 5 amostras em análise terminaram na vizinhança de estruturas secundárias. A sua estabilidade termodinâmica deverá, provavelmente, e aquando da etapa de hibridação no processo de amplificação cíclica, ter impedido a progressão da *Taq* polimerase ao longo das moléculas de DNA matriz. Pelo contrário, relativamente à amostra 4088, a única para a qual foi possível obter uma sequência completa com o tamanho previsto, o resultado obtido poderá significar que na respectiva sequência nucleotídica, não ocorreu a formação de estruturas secundárias suficientemente estáveis para terem comprometido a actividade da *Taq* polimerase.

Uma vez que não foi possível dar continuação ao estudo genético da região 5'UTR/E1 (dado o baixo rendimento obtido na sequenciação dos fragmentos resultantes da amplificação desta região, traduzido pela obtenção de apenas uma sequência completa de entre 6 amplicões analisados), a genotipagem e análise filogenética destas amostras foi restrita ao estudo das regiões genómicas definidas como E1/E2 e NS5A/NS5B.

A genotipagem por análise filogenética é um método robusto que permite a identificação de agrupamentos monofiléticos, onde as sequências em análise segregam com sequências de referência, previamente caracterizadas, e que, neste caso, incluíram representantes dos 6 genótipos de GBV-C. O estudo das regiões codificantes do genoma de GBV-C, definidas como E1/E2 e NS5A/NS5B, foi realizado a partir de amplicões obtidos por *nested*-PCR a partir de 45 e 41 amostras de cDNA viral, respectivamente. A metodologia utilizada, após optimização das condições experimentais de amplificação permitiu, na maioria dos casos, a obtenção de fragmentos com elevada especificidade. No entanto, relativamente à região NS5A/NS5B, não foi possível a obtenção de um produto de amplificação para 4 das amostras biológicas em análise (3638, 6004, 7888 e 8731), ainda que tenham sido testadas alternativas ao protocolo de amplificação. Tal poderá ser sido consequência de degradação do RNA viral como resultado de

repetidas etapas de congelamento/descongelamento das amostras analisadas. À semelhança do que foi anteriormente já mencionado, a possível existência de polimorfismos nas sequências nucleotídicas em questão poderá ter condicionado a hibridação dos *primers* usados com as mesmas, apesar de terem sido utilizados *primers* degenerados, com o intuito de abranger, tanto quanto possível, a diversidade genética viral.

Após amplificação, os produtos de PCR obtidos para as duas regiões (E1/E2, NS5A/NS5B) foram sequenciados em ambas as cadeias, de forma a minimizar a eventual introdução de erros provenientes do processo de sequenciação. As sequências resultantes são sobreponíveis permitindo a construção de sequências consenso. Devido à baixa qualidade obtida como resultado da sequenciação directa dos amplicões obtidos para 10 das amostras analisadas (2 referentes à região E1/E2 e 8 referentes a NS5A/NS5B), possivelmente devido à presença de produtos inespecíficos que resultaram do processo de amplificação, os fragmentos correspondentes a essas amostras foram novamente amplificados e clonados num vector plasmídico. A análise da sequência das moléculas de DNA recombinante contendo os insertos desejados permitiu a obtenção de sequências de qualidade evidente, e a persecução da análise pretendida.

A caracterização genética das sequências de GBV-C portuguesas, teve por base uma análise filogenética que compreendeu a construção de filogramas (árvores filogenéticas em que o comprimento dos ramos é função da distância genética) de acordo com três modelos de reconstrução filogenética: método de junção-de-vizinhos (*Neighbor-Joining* - NJ), análise Bayesiana com base no processo aleatório de cadeias de Monte-Carlo (*Markov Chain Monte Carlo* – MCMC) e pesquisa de árvores filogenética por aplicação do critério de Máxima Verosimilhança (*Maximum Likelihood* – ML).

As árvores filogenéticas surgem como um suporte gráfico natural das filogenias, permitindo representar os agrupamentos dos *taxa* em estudo e, simultaneamente, indicar os diferentes graus de proximidade filogenética existentes entre eles. No caso da análise de sequências de DNA, esta pressupõe que as posições nucleotídicas comparadas sejam ocupadas por nucleótidos homólogos, divergentes de um ancestral comum. A análise da topologia das árvores prevê a identificação de grupos monofiléticos, nos quais estão contidos os *taxa*, e todos os seus descendentes, os quais se encontram agrupados pela partilha de um ancestral comum, não partilhado por mais nenhum *taxa* destes excluídos.

No que diz respeito à região E1/E2, a atribuição de um genótipo a cada uma das 45 sequências de GBV-C em estudo resultou da sua distribuição por árvores filogenéticas com topologias essencialmente semelhantes. Em cada uma delas verificou-se uma clara segregação de 10 estirpes de GBV-C com sequências de referência correspondentes ao genótipo 1, e 34 com as sequências do genótipo 2. A formação destes grupos monofiléticos foi suportada quer por valores de *bootstrap* superiores a 75% (análise pelo método de junção-de-vizinhos) quer por valores de probabilidade *a posteriori* de 1,00 (análise *Bayesiana*). Os resultados não foram, de forma alguma, inesperados. De facto, dada a localização geográfica de Portugal no espaço europeu, e tendo em conta estudos anteriores (Naito e Abe, 2001; Hattori *et al.*, 2003; Muerhoff *et al.*, 2003), era de prever que a grande maioria das estirpes virais analisadas viessem a ser associadas ao genótipo 2. No entanto, o grande grupo monofilético que nas árvores obtidas define o genótipo 2 de GBV-C revelou uma elevada heterogeneidade, quando considerada a origem geográfica das sequências virais analisadas. A sua análise reflectiu, assim, esta clara dispersão geográfica, com aproximadamente 56% (19/34) das sequências a terem origem em países tão distintos como a Bolívia, a Jamaica, os EUA, o Japão, a China, a Alemanha, a Grécia, a Inglaterra, ou o Egipto. Tal facto pode ser, possivelmente, justificado não só como consequência do estabelecimento de prováveis redes de transmissão do vírus, tais como as que foram já caracterizadas para o HCV (Laar *et al.*, 2009), como ainda pelos fluxos migratórios inerentes ao processo de globalização.

No mesmo sentido, as estreitas relações mantidas entre Portugal e os Países Africanos de Língua Oficial Portuguesa que ainda hoje suportam a emigração de cidadãos provenientes das antigas colónias portuguesas no continente africano, contribuíram para um elevado número de indivíduos de etnia africana residentes no nosso país. Considerando que a maioria das estirpes de GBV-C circulantes em África está associada ao genótipo 1 (Xiang *et al.*, 2005; Li^b *et al.*, 2006), e dada a elevada percentagem de indivíduos com etnia africana incluídos neste estudo (28,50%), não foi, mais uma vez, estranha, a observação de que 8 (80,00%) das 10 sequências de GBV-C que agruparam com referências do genótipo 1, tenham sido amplificadas a partir de amostras de RNA viral extraído de plasma proveniente de indivíduos com nacionalidade/etnia africana. De facto, a análise estatística dos resultados obtidos revelou uma associação estatisticamente significativa entre as amostras de GBV-C do genótipo 1 e a etnia Africana dos indivíduos incluídos neste estudo. A caracterização genética deste grupo de sequências classificadas como pertencentes ao genótipo 1 poderá ser relevante numa perspectiva futura, na medida em que

poderá constituir um possível foco de dispersão de variantes destas estirpes de GBV-C pela Europa Ocidental.

Um processo evolutivo fundamental que tem sido subestimado nos estudos envolvendo a caracterização de estirpes de GBV-C, e que poderá ter um papel crucial na determinação de distintos padrões de diversidade genética, é o da recombinação. Sabendo à partida que a recombinação homóloga é uma conhecida característica dos Flavivírus, Worobey e Holmes demonstraram que este fenómeno afecta a evolução do GBV-C, contribuindo para modelar a diversidade genética deste vírus (Worobey e Holmes, 2001). A interpretação da análise filogenética decorrente do presente estudo demonstrou a presença de uma sequência recombinante, associada à amostra 6388. Esta sequência segregou nas árvores filogenéticas das restantes do genótipo 1, sugerindo uma natureza divergente das demais sequências. Deste modo, foi avaliada, através de uma análise de *bootscanning*, a possibilidade de que a ocorrência de recombinação genética estivesse na génese desta observação (Figura 3.17). Os resultados obtidos evidenciaram uma associação filogenética clara da parte central da sequência analisada com sequências de referência de genótipo 1 (com elevado valor de *bootstrap*, 85%) revelando, desta forma, um fenómeno de recombinação intragenotípico. Contudo, as extremidades 3' e 5' não foram associadas com nenhuma das referências utilizadas no alinhamento múltiplo que serviu de base a esta análise. Tal facto pode justificar-se pela subrepresentação de alguns genótipos, nomeadamente no que respeita aos genótipos 4 e 6, em termos de número de sequências disponíveis para análise, e/ou pela carência de sequências virais provenientes de habitantes de regiões geográficas próximas de Portugal.

Ainda que a atribuição de um genótipo às restantes 34 sequências de GBV-C em estudo tenha resultado da sua associação a um grupo monofilético reconhecido como genótipo 2 (G2), foi evidente a formação, no seu conjunto, de dois subgrupos, designados por G2^a (n=9) e G2* (n=16), aparentemente afastados das restantes sequências, agrupadas em G2 (n=10). No entanto, a consistência do subgrupo G2^a foi apenas suportada pela análise Bayesiana efectuada, não tendo sido associados valores de *bootstrap* significativos a este sub-agrupamento pela análise de NJ.

Este facto revelou-se de acordo com a análise gráfica obtida pela ferramenta PCOORD (Figura 3.16), onde a dispersão das sequências em estudo indicou uma clara distinção entre aquelas que segregaram por G1, G2 e G2*, sem que a formação de um grupo distinto relativo a G2^a tenha sido observada. Para testar a hipótese de que G2* correspondesse a um novo genótipo viral, foram calculadas as distâncias genéticas das sequências de um mesmo genótipo (distâncias

intra-genotípicas) ou de genótipos distintos (distâncias intergenotípicas) e comparadas entre si. Os valores obtidos foram expressos, graficamente, num histograma (Figura 3.18 A). As distribuições auferidas foram, de seguida, confrontadas com os valores resultantes de uma análise semelhante mas restrita a G2* vs G2 (por este ser o grupo filogeneticamente mais próximo de G2*). Apesar da óbvia segregação G2* enquanto grupo monofilético que diverge, nas árvores filogenéticas obtidas, das demais sequências agrupadas no genótipo 2, a distribuição das distâncias genéticas relativas às comparações de sequências G2* e G2 recaiu na região mediana que separa as distribuições de distâncias intra e intergenotípicas. Este resultado, complexo pelo facto de que a análise efectuada foi restrita, apenas, a uma secção, e não à totalidade, do genoma viral, não apoiou a hipótese inicial de que G2* viesse a constituir um novo genótipo de GBV-C. Este resultado surge, provavelmente, devido a efeitos fundadores que podem ser consequência da introdução de um número limitado de estirpes virais numa área geográfica singular e/ou num grupo de indivíduos incluídos em grupos com características epidemiológicas semelhantes (como o são os UDIs). O seu isolamento genético, associado ao aumento da divergência genética, poderá conduzir estes grupos fundadores à constituição de novos genótipos virais. Embora este possa ser o trajecto que levou à origem de G2*, a verdade é que, e tanto quanto os dados obtidos o permitem indicar, a sua identidade genética não é, ainda, suficiente para que formem um novo genótipo viral.

A existência de sub-divisões dentro dos grupos identificados como os genótipos reconhecidos de GBV-C, designados subtipos, tem sido apontada por vários autores, em particular no que diz respeito aos subtipos dos genótipos 1, 2 e 5 (Muerhoff *et al.*, 1996; Liu *et al.*, 2003; Muerhoff *et al.*, 2005). Neste sentido, a análise de sequências virais classificadas no genótipo 2 levou alguns autores a sugerir a existência de dois subtipos, *a* e *b*, tal como mencionado anteriormente. Os subtipos pertencentes ao genótipo 2, até à data reportados, corroboram a heterogeneidade genética deste agrupamento (Muerhoff *et al.*, 1996; Muerhoff *et al.*, 1997; Smith *et al.*, 1997; Lopez-Alcorocho *et al.*, 1998; Muerhoff *et al.*, 2003; Castelain *et al.*, 2004; Hekmat *et al.*, 2008). No entanto, não foi possível comparar directamente G2* com esses subtipos dado que a análise dos mesmos foi efectuada, na sua maioria, a partir de secções dos genoma correspondentes à região 5'UTR as quais, não sendo sobreponíveis com a região E1/E2 aqui analisada, impossibilitou a sua comparação directa. Para além disso, alguns estudos não são claros quanto à identificação de subtipos dentro de genótipos uma vez que, frequentemente, estes agrupamentos

não são suportados, nas árvores filogenéticas, por valores de *bootstrap* significativos (Muerhoff *et al.*, 1997; Smith *et al.*, 1997; Hekmat *et al.*, 2008).

Com base nos resultados da análise filogenética obtidos, foi testada a eventual dependência entre as variáveis de carácter epidemiológico (relativas às amostras em estudo) e a sua associação a determinado genótipo, através do teste de χ^2 (com correcção de *Yates*). O cálculo estatístico mostrou uma independência da maioria (excepção para a já referida associação entre o genótipo 1 com indivíduos de etnia africana) das variáveis estudadas ($P > 0,05$). Mais uma vez é de ressaltar que estes resultados poderão estar condicionados a uma amostragem relativamente reduzida, bem como à escassa informação epidemiológica relativa aos indivíduos co-infectados.

A caracterização genética da região E1/E2 das estirpes de GBV-C identificadas, foi complementada com uma análise de substituições nucleotídicas. A reduzida taxa de substituições nucleotídicas de GBV-C (estimada em $9,0 \times 10^{-6}$ por sítio, por ano), é caracterizada por um excesso de substituições sinónimas, a qual se traduz no elevado grau de conservação da sequência poliproteica (94% de identidade), mesmo quando são comparadas estirpes de regiões geográficas afastadas (Pavesi, 2001). A pressão selectiva exercida sobre as 45 sequências de GBV-C em análise foi concordante com a tendência evolutiva para uma conservação a nível primário da poliproteína viral, largamente descrita em estudos anteriores (Wang *et al.*, 1997; Suzuki *et al.*, 1999; Simmonds e Smith, 1999). De facto, a média obtida pela razão $d_s/d_N = 25,8$, reflecte a acumulação de substituições sinónimas nestas sequências e, consequentemente, um reduzido grau de variabilidade da sequência dos potenciais produtos do genoma viral, também constatado no cálculo de d_s/d_N por codão. Por outro lado, e contrariando a observação anterior, na análise das 35 sequências que segregaram com aquelas definidas como pertencentes ao genótipo 2 ($G2^* + G2^a + G2$) foi demonstrada uma pressão positiva exercida sobre as mesmas, com $d_s/d_N = 0,92$ evidenciando como tal, uma tendência para ocorrência de substituições não sinónimas. Uma vez que a região genómica analisada codifica as glicoproteínas do GBV-C, e estando estas localizadas à superfície do virião, a ocorrência de substituições não sinónimas nesta região do genoma poderá ser o reflexo de escape do vírus à resposta imunitária do hospedeiro.

O grau de variabilidade relativo aos resíduos de aminoácidos das sequências agrupadas no genótipo 2 foi ainda completado através do cálculo da Entropia de Shannon. Tendo em conta o

carácter divergente das sequências G2*, a análise efectuada baseou-se na comparação das sequências que o integram relativamente às demais sequências G2 analisadas (*background*). Os valores obtidos revelaram um elevado grau de conservação em G2* de determinados resíduos de aminoácidos na sequência primária dos produtos codificados pela região analisada ($P < 0,005$). São disto exemplos os resíduos de alanina (A58), serina (S110), glutamina (Q134) e fenilalanina (F182) destacados, a vermelho, na Figura 3.21. Analogamente, o resíduo de asparagina N86 encontrou-se mais conservado no grupo que constituiu o respectivo *background*. Este resultado evidencia posições que poderão revelar uma eventual importância biológica e, como tal, servirem como ponto de partida para estudos mais aprofundados que o permitam esclarecer.

Discordante das relações filogenéticas estabelecidas pelo estudo da região E1/E2, a análise filogenética obtida com as sequências correspondentes à região subgenómica NS5A/NS5B evidenciou alguns resultados ambíguos face ao estudo da primeira região. Enquanto a segregação das sequências E1/E2 ocorreu com formação de grandes grupos monofiléticos consistentes, representando estes os genótipos de GBV-C (independentemente do método de inferência filogenética utilizado), o mesmo não sucedeu com parte das sequências correspondentes à região NS5A/NS5B. No entanto, por análise Bayesiana, um algoritmo matemático com grau de complexidade superior ao da análise por NJ, e ainda pelo facto de conferir probabilidade aos agrupamentos identificados nas árvores filogenéticas, a consistência dos resultados de genotipagem obtidos para esta região foi verificada examinando, apenas, as árvores obtidas com este método de reconstrução filogenética. Esta escolha foi ainda sustentada pelo facto de os três modelos de reconstrução filogenética utilizados terem recriado árvores com uma topologia essencialmente idêntica entre si.

A constatação da dispersão de algumas das sequências NS5A/NS5B foi revelada, particularmente, pelas sequências identificadas como 8227NS5B, 1641NS5B, 0594NS5B, 1631NS5B, 9317NS5B e 6388NS5B. Relativamente às três primeiras não houve segregação com nenhum agrupamento específico, embora se encontrem filogeneticamente próximas de sequências de referência de genótipo 1. No caso da sequência 8227NS5B, agrupada no genótipo 2 pela análise da região E1/E2, segregou com a referência AB003291 (G1) que, curiosamente, diverge de modo singular das demais sequências em análise. Por seu lado as sequências 1641NS5B e 0594NS5B constituíram um grupo que incluiu AB013500 (G1) e AB003292, uma sequência separada precocemente deste grupo e classificada dentro do genótipo 6 (Muerhoff *et*

al., 2006). Estes três casos evidenciaram não só incongruência com os resultados obtidos no decurso da análise efectuada para a região E1/E2, como também demonstraram a não formação de grupos consistentes. Disto é exemplo a formação de um agrupamento monofilético que incluiu uma terceira sequência de referência G1 (U36380, com origem no Gana) com a qual 6388NS5B segrega precocemente das sequências 9317NS5B, 9584NS5B, 8395NS5B, 6073NS5B, 0123NS5B e 265NS5B, e que, juntas formam um agrupamento suportado com uma probabilidade *à posteriori* de 0,92. Contudo, o fragmento E1/E2 relativo a esta amostra foi agrupado, claramente, com sequências de genótipo 2, revelando mais uma vez, o desacordo nos resultados obtidos pela análise das duas regiões subgenómicas. Por fim, a sequência 1631NS5B aparenta, igualmente, um comportamento distinto ao divergir prematuramente do grupo identificado por G2^a.

As dificuldades associadas à genotipagem da região NS5A/NS5B amplificada das amostras 1641, 8227, 1631 e 6388 justificou a pesquisa da sua potencial natureza recombinante, facto esse que, por si só, poderia explicar os resultados obtidos por análise filogenética. De facto, a detecção da presença de eventos de recombinação, efectuada através da ferramenta *Simplot*, foi demonstrada relativamente a todas elas. Na primeira, a extremidade 5' apresentou homologia com sequências de genótipo 1, enquanto que na região a montante da extremidade 3' esta foi partilhada com uma sequência correspondente ao genótipo 6. No entanto, apesar da ocorrência de recombinação entre sequências de genótipos diferentes, a atribuição, de forma não ambígua, de um genótipo a determinadas porções da sequência alvo não foi possível, impossibilitando a identificação dos pontos de recombinação. Este resultado corrobora então a posição desta sequência nas árvores filogenéticas construídas, denotando-se um grau de proximidade com sequências quer de genótipo 1 (Gana) quer de genótipo 6 (Indonésia). Relativamente à sequência 8227NS5B, apesar de grande parte desta estar mais próxima do ponto de vista evolutivo com sequências de genótipo 1, foi revelada uma estrutura em mosaico com uma referência de genótipo 5 manifestando, como tal, uma presumível natureza recombinante. No entanto, a origem genética da extremidade 3' permaneceu desconhecida por não ter revelado homologia significativo com nenhuma das sequências de referência utilizadas. De forma semelhante, a sequência analisada como 6388NS5B demonstrou uma estrutura em mosaico com regiões recombinantes na extremidade 5' entre sequências de genótipo 1 e 3, na zona mediana, entre representantes dos genótipos 3 e 2, e na zona terminal 3', com o genótipo 1. Por último, a sequência correspondente à amostra a 1631 não evidenciou um grau de recombinação significativo, uma vez que toda a sequência nucleotídica

estabeleceu elevada identidade com referências do genótipo 1, embora numa região de aproximadamente 200nt, na sua secção média, se verifique uma diminuição de homologia com essas referências, o que suporta o facto da divergência não justificada desta sequência das restantes de genótipo 2 aquando análise filogenética. Em suma, a caracterização destas sequências como hipotéticos recombinantes poderá então, no que respeita à região NS5A/NS5B, justificar a ausência de uma segregação consistente de algumas das sequências analisadas com os grupos monofiléticos formados por reconstrução de árvores filogenéticas. No entanto, a análise de recombinantes exposta pode ter sido condicionada pelo reduzido número de referências disponíveis, respectivamente no que diz respeito aos genótipos 4, 5 e 6.

Por outro lado, as observações relativas à análise filogenética para as sequências 1631NS5B, 1641NS5B, 0594NS5B, 6388NS5B e 8227NS5B foram ainda corroboradas pelo cálculo das distâncias genéticas entre cada uma delas e os agrupamentos G1, G2, G2^a e G3 individualmente (Tabela 3.3 B). Curiosamente, a sequência 6388NS5B encontra-se geneticamente mais próxima de G3 do que das sequências que segregaram com G1, facto também evidenciado na árvore filogenética obtida pelo método de junção-de-vizinhos.

A inconsistência das relações filogenéticas estabelecidas pela região NS5A/NS5B contrasta com a análise efectuada relativamente à região codificante E1/E2, já anteriormente evidenciada por Smith e colaboradores (Smith *et al.*, 2000). Alguns autores sugerem que esta diferença possa ser justificada pela ocorrência de mecanismos evolutivos discrepantes nas várias regiões do genoma de GBV-C ou que estes as afectem de modo diferencial. Desta forma, a análise combinada destas regiões poderia produzir uma reconstrução filogenética com menor nível de confiança face ao estudo individualizado dessas mesmas regiões (Smith *et al.*, 2000). No entanto, a validade do estudo baseado em sequências do genoma completo é suportada pela correlação entre as relações filogenéticas resultantes e a origem geográfica. Uma explicação alternativa advém do elevado grau de conservação da sequência aminoacídica que constitui a poliproteína de GBV-C ($d_S/d_N=30,30$ com uma divergência $<11\%$) comparativamente com os subtipos de HCV ($d_S/d_N=10,60$ com uma divergência $<10\%$). Os grupos filogenéticos de GBV-C resultam, como tal, não só da ocorrência de substituições nucleotídicas de tipo sinónimo, como da existência de locais inalterados, possivelmente devido aos entraves impostos pela formação de estruturas secundárias, já demonstradas nas regiões NS5A e NS5B (Simmonds e Smith, 1999; Cuceanu *et al.*, 2001).

Consequentemente, as estirpes de GVB-C com uma origem evolutiva comum, podem partilhar apenas um pequeno número de polimorfismos e daí a análise das regiões subgenómicas reverter na produção de árvores filogenéticas incongruentes. Por outro lado, a consistência das análises filogenéticas obtidas pelo estudo do gene E2 e do genoma viral completo pode estar associada ao facto desta ser a região com maior grau de variabilidade na sequência nucleotídica de GBV-C (Katayama *et al.*, 1998; Erker^b *et al.*, 1996). Assim, as relações filogenéticas estabelecidas pela análise das sequências de genoma completo podem depender, em grande parte, das substituições que ocorrem em E2 (Smith *et al.*, 2000). No entanto, contrariando esta hipótese, a análise filogenética da região 5'UTR de GBV-C tem sido também traduzida de forma fidedigna e consistente com as filogenias recriadas com base na análise da totalidade do genoma viral (Sathar e York, 2001; Muerhoff *et al.*, 2005).

A caracterização das sequências da região NS5A/NS5B foi assim complementada com o cálculo das razões entre taxas de substituições sinónimas e não-sinónimas (d_S/d_N) em dois grupos de sequências distintos. Em ambos, o conjunto caracterizado pelas 36 sequências onde foram obtidos resultados não ambíguos pela análise filogenética e o grupo de 30 amostras relativo ao genótipo 2 (incluiu as amostras que segregaram com G2, G2* e G2^a), foi observada uma tendência para a variabilidade das sequências nucleotídicas com valores de $d_S/d_N < 1$. Consequentemente, e de forma semelhante à análise obtida relativamente ao conjunto de sequências G2 respectivo à região E1/E2, as amostras em estudo encontram-se aparentemente sobre uma pressão positiva, com um acumular de substituições não sinónimas superior ao das substituições aminoacídicas silenciosas. Este conspecto sugere e confirma a diversidade genética existente na região NS5A/NS5B de GBV-C anteriormente demonstrada por Pavesi (Pavesi, 2001).

No entanto, as relações filogenéticas não ambíguas das restantes 30 estirpes de GBV-C em estudo estabeleceram grupos monofiléticos com elevado suporte estatístico ($P > 0,9$) onde 6 sequências foram agrupadas com referências de genótipo 2, 9 foram distribuídas por um grupo que divergiu prematuramente deste último e que, por analogia com o estudo prévio da região E1/E2, foi denominado por G2^a, enquanto que 15 segregaram num putativo subgrupo G2*, já distinguido pela análise dos fragmentos de GBV-C correspondentes à região E1/E2.

À semelhança do estudo efectuado relativamente à região E1/E2, a hipótese da formação de dois putativos subgrupos G2^a e G2* foi testada através da utilização da ferramenta PCOORD e pela análise comparativa das distâncias genéticas considerando as possíveis partições do genótipo 2 (G2, G2^a e G2*). Ambas revelaram a proximidade genética entre as sequências do grupo G2 com

as do grupo G2^a, não permitindo a classificação deste último como um subgrupo distinto, apesar da evidente divergência prematura deste conjunto de sequências com aquelas representadas por G2 nas 3 árvores filogenéticas. Relativamente ao subgrupo G2*, e analogamente ao sucedido na análise da região E1/E2, foi demonstrada uma clara divergência dos demais analisados, facto concordante com a precedente análise filogenética.

Este facto foi ainda corroborado pelo cálculo das distâncias genéticas entre G1, G2 (incluindo as sequências G2^a) e G2* (Tabela 3.3 A) tendo-se verificado que os grupos G2* e G1 são aqueles que apresentam maior distância genética entre si (12,4%). Tal como esperado, a distância genética dentro de cada respectivo grupo em análise (G2, G2* e G2^a) revelou ainda o grupo G2 como o mais heterogéneo de entre aqueles em estudo (considerando que as referências utilizadas, para além das relativas a Lisboa, têm origens geográficas bastante distintas), com uma percentagem de 9,5%. Interessante foi a distância média de 6,9% entre as sequências do subgrupo G2^a que evidenciaram este conjunto como sendo o mais homogéneo.

Deste modo, a análise dos fragmentos de 15 amostras comuns para ambas as regiões genómicas estudadas evidenciou a segregação exclusiva de um grupo monofilético específico (designado como G2*), o qual pelas várias análises efectuadas quer relativamente à região E1/E2, quer ao estudo da região NS5A/NS5B, poderá representar um putativo subtipo de estirpes virais do genótipo 2, possivelmente associado à área metropolitana de Lisboa.

Por analogia com o estudo feito relativamente à região genómica E1/E2, o grau de variabilidade das posições dos aminoácidos nas sequências proteicas relativas aos fragmentos da região NS5A/NS5B em estudo foi ainda estimada de acordo com a função da Entropia de Shannon. Como grupo alvo (*query*) foram seleccionadas as sequências pertencentes a G2*, por terem segregado de forma particular das remanescentes de genótipo 2. Para efeitos de comparação (*background*) seleccionaram-se, sequências de GBV-C pertencentes ao genótipo 2 (incluindo as sequências G2 obtidas neste estudo e um grupo de sequências de referência do mesmo genótipo). Foi denotado um elevado grau de conservação relativamente ao resíduo de leucina, L51 (Figura 3.29) no grupo G2*, com valores estatisticamente significativos ($P < 0,005$) revelando uma caracterização genética distinta do grupo filogeneticamente mais próximo (G2).

Tendo sido já estabelecida uma relação entre a carga viral de GBV-C e a eficácia da terapia por interferão em doentes com doença hepática por HCV (Takahashi *et al.*, 1997), a influência das alterações em determinadas sequências de aminoácidos de regiões genómicas do GBV-C tem

sido explorada (Fujisawa *et al.*, 2000). A região NS5A de GBV-C apresenta, por analogia com a mesma região genómica do HCV, uma sequência homóloga com a região determinante de sensibilidade ao interferão (ISDR) do HCV (Orito *et al.* 1997), pelo que Fujisawa e colaboradores estudaram a interferência das posições na sequência proteica desta região relativamente à eficácia da terapia com interferão em doentes com hepatite C (Fujisawa *et al.*, 2000 109). Mais recentemente, estudos relativos à inibição da replicação do HIV por determinados domínios peptídicos do fragmento genómico NS5A do GBV-C, têm também sido descritos (Xiang *et al.*, 2006; Chang *et al.* 2007; Xiang *et al.*, 2008). Apesar do presente estudo incidir maioritariamente no estudo da região NS5B, é de realçar a importância das relações biológicas que podem surgir pela variabilidade de determinadas posições aminoacídicas em regiões particulares do genoma viral.

ENSAIOS DE PCR EM TEMPO REAL - QUANTIFICAÇÃO RELATIVA DA CARGA VIRAL DE GBV-C

Do ponto de vista clínico, a quantificação da virémia de um vírus poderá ter um impacto directo no diagnóstico e/ou progressão da doença, determinando ainda o início da terapêutica prescrita. Esta situação é claramente evidente, e especificamente relevante, para o HIV (Lima *et al.*, 2009). Com um carácter biológico que reflecte um elevado grau de variabilidade, e onde a formação de quasiespécies é comum, tornou-se necessário a aplicação de métodos quantitativos suficientemente sensíveis e específicos para uma determinação exacta da carga viral deste vírus, nas quais a diversidade genética viral viesse a ter o menor impacto possível (Buonaguro *et al.*, 2007; Mendonza e Soriano, 2009). Dada a já discutida possibilidade da existência de uma relação benéfica que advém da co-infecção entre o HIV e o GBV-C, a quantificação da carga viral pode ter um papel fundamental na compreensão dos eventuais mecanismos envolvidos na dinâmica da interacção entre estes dois vírus (Lefrère^b *et al.*, 1999; Li^a *et al.*, 2006). Recentemente, Maidana-Giret e colaboradores sugeriram que a associação entre a replicação do GBV-C e uma baixa activação das células T poderá ser o mecanismo chave envolvido na protecção conferida por este vírus na progressão para a doença nos indivíduos infectados pelo HIV (Maidana-Giret *et al.*, 2009).

À semelhança do que acontece para uma grande variedade de outros vírus, um dos métodos mais comumente usados na quantificação da virémia de GBV-C baseia-se na PCR em Tempo Real. Esta metodologia combina os processos de amplificação e de detecção simultaneamente, na medida em que a fluorescência emitida por fluoróforos no meio de reacção é registada, em

tempo real, à medida que os produtos são sintetizados (Kubista *et al.*, 2006). A monitorização do ensaio ocorre durante todo o processo de produção de amplicões, e a cada ciclo de PCR, produzindo uma curva de amplificação de acordo com a concentração do produto sintetizado e em função do número de ciclos (Mackay *et al.*, 2002; Kubista *et al.*, 2006). Uma das principais vantagens, para além do permitir seguir a reacção em tempo real, é o facto de a detecção e/ou quantificação por PCR em Tempo Real eliminar os processos pós-amplificação (ex: electroforese em gel de agarose) reduzindo as manipulações a efectuar em etapas, *a priori*, desnecessárias. Por outro lado, a reprodutibilidade deste método está associada à aquisição dos dados de quantificação na fase exponencial inicial da PCR, na qual a sua variabilidade é menor que nas leituras dos ciclos finais. É ainda um sistema altamente sensível que permite detectar quantidades mínimas de moléculas de DNA numa amostra biológica (Kubista *et al.*, 2006; Ratcliff *et al.*, 2007).

São vários os sistemas de detecção usados nesta técnica, embora a quantificação da carga viral de GBV-C, por utilização de sondas *TaqMan*, tenha sido o método eleito pela maioria dos autores (Hattori *et al.*, 2003; Castelain *et al.*, 2004; Ruiz *et al.*, 2006). O método com sondas *TaqMan* é mais específico pela necessidade de fazer hibridar às moléculas matriciais três oligonucleotídeos distintos (2 *primers* de amplificação e 1 sonda de detecção) mas por outro lado implica um custo mais elevado pela utilização de um oligonucleotídeo/sonda modificado(a) com fluorocromos (Kubista *et al.*, 2006). No entanto, esta elevada especificidade pode, por vezes, condicionar a reacção de amplificação em situações sub-óptimas, por exemplo no caso de polimorfismos entre a sequência alvo a amplificar e a sequência nucleotídica da sonda, impossibilitando a hibridação entre ambos.

Dada a importância biológica associada à carga viral, foi então proposto não só quantificar a virémia de GBV-C em 9 amostras biológicas escolhidas aleatoriamente, como também comparar a reprodutibilidade obtida com duas técnicas distintas de detecção de fluorescência: a utilização do corante *SYBR Green I* e de uma sonda de hidrólise (sonda *TaqMan*) desenhada por Hattori e colaboradores (Hattori *et al.*, 2003). A construção de uma recta padrão foi efectuada para ambos os sistemas de detecção de sinal utilizados. A divergência de valores obtidos, apesar da utilização do mesmo DNA matriz, a mesma região alvo (5'UTR) e condições de reacção idênticas, foi evidente pela não sobreposição ou paralelismo das rectas derivadas. O método de detecção baseado na utilização de *SYBR Green I* depende da fluorescência por este emitida à medida que se intercala no DNA em cadeia dupla gerado. Neste sentido, os ensaios que utilizam

este agente podem revelar resultados dúbios por ligação do corante a moléculas de DNA amplificadas de forma inespecífica, ou a moléculas de ácidos nucleicos que, mesmo em cadeia simples, tendam a formar estruturas secundárias. De igual forma, verificou-se ainda uma diminuição na sensibilidade de detecção de sinal nos ensaios com sonda *TaqMan* (máximo de detecção até uma diluição de 10^{-9}) relativamente aos efectuados com o corante intercalante *SYBR Green I* (detecção até uma diluição de 10^{-11}). Esta observação está em concordância com o provável aumento de especificidade resultante com o primeiro método (sonda *TaqMan*) face ao último (corante *SYBR Green I*) e, consequentemente, detecção de uma menor quantidade de moléculas. No entanto, a consistência de resultados obtidos entre os dois sistemas de detecção tem sido comprovada (Arikawa *et al.*, 2008).

Ainda que a análise semi-quantitativa tenha sido efectuada com um número limitado de amostras, a partir da regressões lineares das duas rectas padrão, e com base nos valores de Cts auferidos no decurso desta análise, foi possível calcular o número de cópias de DNA por mistura de reacção através de cada um dos sistemas de detecção de fluorescência utilizados. Assim, devido à ausência de consonância no declive de ambas as rectas obtidas (os valores de Ct revelados nos ensaios com a sonda *TaqMan*, foram superiores, facto que resultou num declive igualmente acentuado na recta padrão derivada, e consequentemente, superior àquele auferido pelos ensaios com *SYBR Green I*), a análise com *SYBR Green I*, registou um número de cópias de DNA/ml 100X inferior ao detectado pela sonda *TaqMan*. A gama de valores auferidos por este segundo método, a variar entre $6,03 \times 10^4$ e $1,56 \times 10^7$ cópias/ml é no entanto coerente com estudos anteriores relativos à quantificação da carga viral de GBV-C por recurso a sondas *TaqMan* (Hattori *et al.*, 2003; Castelain *et al.*, 2004). Contudo, e inesperadamente, nas reacções de PCR em Tempo Real relativas às amostras biológicas A, B (no método com sonda *TaqMan*) e C (em ambos os métodos) não foi detectado sinal de amplificação. As três, juntamente com a amostra H, foram tentativamente re-amplificadas por *nested*-PCR, sem formação de amplicões relativamente à amostra C (possivelmente por degradação do DNA matriz, uma vez que esta amostra foi sequenciada relativamente aos fragmentos amplificados correspondentes às regiões E2 e NS5B). O produto amplificado para as restantes três amostras foi sequenciado e analisado por comparação com a sequência da sonda *TaqMan* utilizada. Com base na referida análise, foram identificadas cinco alterações em relação à sequência nucleotídica da sonda (*mismatch*) coincidentes nas amostras A e B (Figura 3.31), facto que justifica a não detecção destas amostras pela utilização da sonda *TaqMan*, inversamente ao sucedido com o corante *SYBR Green I*. A

amostra H, tal como previsto, não revelou a presença de qualquer posição nucleotídica discordante com a sequência respectiva à sonda, o que sustenta o argumento referenciado anteriormente. Deste modo apesar da elevada especificidade demonstrada pelo método de detecção de fluorescência por recurso a sondas de hidrólise, a ocorrência de *mismatches* nas amostras alvo evidencia o quão este tipo de abordagem pode ser afectada pela diversidade genética viral, mesmo quando as quantificações pretendidas o são feitas tendo em conta a amplificação/detecção de amplicões incluídos em regiões conservadas do genoma destes vírus. Este facto é de notória importância dado o elevado grau de variabilidade genética característico dos vírus de RNA. Assim, a comparação de dois métodos distintos de detecção de sinal por PCR em Tempo Real permitiu evidenciar algumas das vantagens e desvantagens da sua aplicação em amostras biológicas, nomeadamente na semi-quantificação da carga viral de GBV-C.

5. CONCLUSÕES E PERSPECTIVAS FUTURAS

No presente estudo foi demonstrada a presença de infecção activa por GBV-C em aproximadamente metade da população amostrada, a qual incluiu mais de duas centenas de indivíduos. O vírus infectou, comumente, indivíduos saudáveis e imunocomprometidos, suspeitando-se que o contágio tenha ocorrido, essencialmente, pelas duas principais vias de transmissão que lhe estão associadas, envolvendo a exposição percutânea (por associação com a elevada taxa de infecção relativa aos UDIs) e a transmissão através de contactos sexuais não protegidos. Apesar da elevada taxa de monoinfecção registada, esta foi superada pela prevalência de GBV-C associada a co-infecções com HIV e HCV. No entanto, deve salientar-se a restrição da população estudada à cidade de Lisboa e seus arredores, bem como a um grupo de amostras de conveniência, já que a amostragem não foi feita tendo em conta a determinação de prevalências de infecção reais. Tal facto limita a interpretação dos valores obtidos, e impede-os de estabelecer qualquer tipo de previsão a nível nacional, pelo que seria interessante estender o trabalho a outras regiões do País.

Ainda que geneticamente próximo do HCV, os dados clínicos mais recentes apontam para a ausência de patogenicidade associada ao GBV-C. A razão para esta “inércia” é ainda desconhecida, mas do ponto de vista evolutivo este vírus pode estar intimamente adaptado à espécie humana (Pavesi, A. 2001). É geralmente aceite que a evolução humana tenha tido início na África ancestral de onde os humanos deverão ter irradiado para fora deste continente (Cann, 2001; Stanyon *et al.*, 2009). A aquisição de sequências virais adicionais, amplificadas a partir de amostras biológicas com origem em populações indígenas em países africanos, bem como na península Arábica, Índia e Ásia, seria uma mais-valia na construção da verdadeira árvore evolutiva, bem como na caracterização da real extensão da diversidade global das variantes de GBV-C. Por outro lado, alguns autores têm sugerido que o baixo nível de variação na sequência aminoacídica observada nas estirpes de GBV-C de distintas regiões geográficas, é indicativa da ausência de diferenças biológicas substanciais entre vírus de genótipos distintos (Simmonds e Smith 1999; Simmonds 2001). Contudo, outros vírus, nomeadamente o HIV, com reduzida variabilidade a nível de composição de aminoácidos, exibem diferenças significativas nos fenótipos resultantes. Este caso é bem explícito relativamente a esse vírus, onde diferentes subtipos com um grau de variabilidade de 10-12% na sequência nucleotídica, apresentam, por exemplo, respostas divergentes quando sujeitos a terapia anti-retroviral (Kantor *et al.*, 2005; Martínez-Cajas *et al.*, 2008 316). Assim, dado o putativo efeito protector que o GBV-C parece

exercer em infecções com HIV (Berzsenyi^a *et al.*, 2005) a caracterização genética deste vírus reveste-se de extrema relevância.

Considerando que, até à data, são desconhecidas quais as variantes de GBV-C em circulação em Portugal, este trabalho permitiu a atribuição de um genótipo a cerca de metade das estirpes identificadas em amostras onde foi detectada virémia de GBV-C. Como esperado, a análise filogenética a partir de um fragmento correspondente, maioritariamente, à região codificante da glicoproteína E2, permitiu o reconhecimento de grupos monofiléticos distintos, com segregação das amostras em estudo com o genótipo 1 e com o genótipo 2 (a maioria). Estes resultados não foram surpreendentes uma vez que estudos prévios demonstraram já a persistência da circulação do genótipo 2 na Europa (Naito e Abe, 2001). A associação de dez amostras com o genótipo 1 (onde 80% correspondem a indivíduos de etnia africana) foi consistente com a presença deste genótipo em países africanos (Xiang *et al.*, 2005 158). Inesperadamente, a identificação de um putativo novo subtipo circulante na zona metropolitana de Lisboa, designado por G2*, comprovou a heterogeneidade genética deste vírus sendo, para além do mais, compatível com a sua associação com uma origem geográfica. Neste sentido, e através da expansão desta análise a uma maior amostra populacional, quer a nível nacional quer com outros países da Europa Ocidental/Mediterrânica, seria de extremo interesse a confirmação deste grupo como um novo subtipo confinado a Lisboa/Portugal ou se, pelo contrário, se encontra em circulação por outras regiões geográficas. Contrariamente, uma análise filogenética da região NS5A/NS5B não revelou uma total congruência com os resultados obtidos com a região E2, em uniformidade com o estudo de Smith e colaboradores (Smith *et al.*, 2000). No entanto, deve registar-se que, também aqui, foi evidente a consistência do grupo monofilético correspondente a G2*.

É difícil chegar a um consenso sobre até que ponto os genótipos de GBV-C influenciam, ou não, a saúde humana, por exemplo, no que diz respeito ao seu impacto na progressão para a doença dos indivíduos co-infectados com HIV ou HCV (Berzsenyi *et al.*, 2009). No entanto, à luz dos prováveis benefícios que advêm de uma infecção com GBV-C em indivíduos seropositivos para HIV, é importante examinar possíveis correlações entre os genótipos de GBV-C e os subtipos de HIV. A determinação de quais os mecanismos moleculares que poderão estar envolvidos na eventual protecção contra a progressão para SIDA em indivíduos co-infectados com GBV-C, requer particular atenção, visto que estes estudos podem, eventualmente, conduzir ao desenvolvimento de novas abordagens terapêuticas e/ou profiláticas que nos permitam limitar o impacto do HIV e/ou do HCV na saúde humana.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAHAM P, 2007. GB virus C/hepatitis G virus - its role in human disease redefined? *Indian Journal of Medical Research*, **125**(6):17-719.
- ADAMS, J. N., PRESCOTT, E. L., JARVIS, M. L., LEWIS, M. C. J., MCCLURE, O. M., SMITH B. D., & SIMMONDS, P., 1998. Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *Journal of General Virology*, **79**(8):1871–1877.
- AGGARWAL, R. & KRAWCZYNSKI, K., 2000. Hepatitis E: An overview and recent advances in clinical and laboratory research. Review. *Journal of Gastroenterology and Hepatology*, **15**(1):9–20.
- ALTER, H. J., 1999. Discovery of non-A, non-B hepatitis and identification of its etiology. *American Journal of Medicine*, **107**(6B):16S–20S.
- ALTER, M. J., GALLAGHER, M., MORRIS, T. T., MOYER, L. A., MEEKS, E. L., KRAWCZYNSKI, K., KIM, J.P. & MARGOLIS, H. S., 1997. Acute non-A-E hepatitis in the United States and the role of hepatitis G virus infection. *New England Journal of Medicine*, **336**(11):741-746.
- ALTER. H. J., NAKATSUJI, Y., MELPOLDER, J., WAGES, J., WESLEY, R., SHIH, J. W., KIM, J. P., 1997. The incidence of transfusion-associated hepatitis G virus infection and its relation to liver disease. *New England Journal of Medicine*, **336**(11):747-754.
- AMARANTE, K. M., MIRANDA C. H., CORAL DE OLIVEIRA, E. C., OLIVEIRA, B. K., WATANABE, E. MA. 2004. Detecção do RNA HGV/GBV-C em Indivíduos Saudáveis não Portadores de HBV, HIV-1/2 e HCV. *Seminário: Ciências Biológicas e da Saúde, Londrina*, **25**, 39-44.
- AN, P., WEI, L., WU, X. Y., YUHKI, N., O'BRIEN, S. J. & WINKLER, C., 1997. Evolutionary analysis of the 5-terminal region of hepatitis G virus isolated from different regions in China. *Journal of General Virology*, **78**(10):2477-2482.
- ARIKAWA, E., SUN, Y., WANG, J., ZHOU, Q., NING, B., DIAL, S.L., GUO, L. & YANG, J., 2008. Cross-platform comparison of SYBR Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics*, **11**, 9:328.
- ASIM, M., POTUKUCHI, K. S., ARORA, A., SINGH, B. & KAR, P., 2008. Hepatitis-G Virus Infection in Multi-transfused Patients and Intravenous Drug Abusers: New Delhi Experience. *Digestive Diseases Sciences*, **53**(5):1383–1389.
- BAGGIO-ZAPPIA, L. G. & GRANATO, H. F. C., 2009. HIV-GB virus C co-infection: an overview. Review. *Clinical Chemistry and Laboratory Medicine*, **47**(1):12–19.

- BARBOSA, J. A., BAGGIO-ZAPPIA, L. G., DOBO, C., ALVES-SOUSA, K. V., LANZARA, A. G., SILVA, G. C. D. I., LANZONI, P. V. & GRANATO, H. F. C., 2009. Analysis of GB virus C infection among HIV-HCV coinfecting patients. *Revista da Sociedade Brasileira de Medicina Tropical*, **42**(5):591-3.
- BARUSRUK, S. & URWIJITAROON, Y., 2006. High prevalence of HGV coinfection with HBV or HCV among northeastern Thai blood donors. *Southeast Asian Journal of Tropical Medicine and Public Health*, **37**(2):289-93.
- BASSIT, L., KLETER, B., RIBEIRO-DOS-SANTOS, G., MAERTENS, G., SABINO, E., CHAMONE, D., QUINT, W. & SAEZ-ALQUEZAR, A., 1998. Hepatitis G virus: prevalence and sequence analysis in blood donors of Sao Paulo, Brazil. *Vox Sanguinis*, **74**(2):83-87.
- BENHAMOU, Y., BOCHET, M., DI MARTINO, V., CHARLOTTE, F., AZRIA, F., COUTELLIER, A., VIDAUD, M., BRICAIRE, F., OPOLON, P., KATLAMA, C. & POYNARD, T. 1999. Liver fibrosis progression in human immunodeficiency virus and hepatitis C virus co-infected patients. *The Multivirc Group. Hepatology*, **30**(4):1054-8.
- BERZSENYI^a M. D., BOWDEN, S. D., BAILEY, J. M., WHITE, C., COGHLAN, P., DUDLEY, J. F. & ROBERTS, K. S., 2005. Male to male sex is associated with a high prevalence of exposure to GB virus C. *Journal of Clinical Virology*, **33**(3):243-246.
- BERZSENYI^b, M.D., BOWDEN, S. D. & ROBERTS, K. S., 2005. GB virus C: Insights into co-infection. Review. *Journal of Clinical Virology*, **33**(4):257-266.
- BERZSENYI, M. D., BOWDEN, S. D., KELLY, A. H., WATSON, M. K., MIJCH, M. A., HAMMOND, A. R., CROWE, M. S. & ROBERTS, K. S., 2007. Reduction in Hepatitis C-Related Liver Disease Associated With GB Virus C in Human Immunodeficiency Virus Coinfection. *Gastroenterology*, **133**(6):1821-1830.
- BERZSENYI, M. D., BOWDEN S. D., ROBERTS, K. S. & REVILL, P., 2009. GB Virus C Genotype 2 Predominance in a Hepatitis C Virus/HIV Infected Population Associated With Reduced Liver Disease. *Journal of Gastroenterology and Hepatology*, **24**(8):1407-1410.
- BIRKENMEYER, G. L., DESAI, M. S., MUEHROFF, S. A., LEARY, P. T., SIMONS, N. J., MONTES, C. C. & MUSHAHWAR K. I., 1998. Isolation of a GB Virus-Related Genome From a Chimpanzee. *Journal of Medical Virology*, **56**(1):44-51.
- BJORKMAN, P., FLAMHOLC, L., NAUCLER, A., MOLNEGREN, V., WALLMARK, E. & WIDELL, A., 2004. GB virus C during the natural course of HIV-1 infection: viremia at diagnosis does not predict mortality. *AIDS*, **18**(6):877-86.
- BJORKMAN, P., NAUCLER, A., WINQVIST, N., MUSHAHWAR, I. & WIDELL, A., 2001. A casecontrol study of the transmission routes for GB virus C/hepatitis G virus in Swedish blood donors asking markers for hepatitis C virus infection. *Vox Sanguinis*, **81**(3):148-53.

- BLUMBERG, B. S., ALTER, H. J. & VISNICH, S., 1965. "A new" antigen in leukemia sera. *Journal of the American Medical Association*, **51**, 191:541-6.
- BOURLET, T., BERTHELOT, P., GRATTARD, F., GENIN, C., LUCHT, R. F. & POZZETTO, B., 2002. Detection of GB virus C/hepatitis G virus in semen and saliva of HIV type-1 infected men. *Clinical Microbiology and Infection*, **8**(6): 352–357.
- BOWDEN, S., 2001. New hepatitis viruses: Contenders and pretenders. Review. *Journal of Gastroenterology and Hepatology*, **16**(2), 124–131.
- BREEN, E. C., 2002. Pro and anti-inflammatory cytokines in human immunodeficiency virus infection and acquired immunodeficiency syndrome. *Clinical Pharmacology & Therapeutics*, **95**: 295–304.
- BRUMME, Z. L., CHAN, K. J., DONG, W. W. Y., MO, T., WYNHOVEN, B., HOGG, R. S., MONTANER, J. S., O'SHAUGHNESSY, M. V. & HARRIGAN, P. R., 2002. No association between GB virus-C viremia and virological or immunological failure after starting initial anti-retroviral therapy. *AIDS*, **16**(14):1929–33.
- BUONAGURO, L., TORNESELLO, M. L. & BUONAGURO, F. M. 2007. Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. Minireview. *Journal of Virology*, **81**(19):10209–10219.
- CANN R. 2001. Genetic Clues to Dispersal in Human Populations: Retracing the Past from the Present. *Science* , **291**(5509):1742-8.
- CAO, M. M., REN, H., ZHAO, P., PAN, W., CHEN, Q. L. & QI, Z. T., 2009. Persistent replication of the GBV-C subgenomic replicons in Huh7 cells. *Journal of Virology Methods*, **157**(2):168-174.
- CARVALHO, A., MARTINHO, A., CIPRIANO, MA., COIMBRA, B. H. & PORTO, A. 1999. Prevalência e Significado Clínico da infecção pelo vírus da “hepatite” G em diversos grupos de doentes. *Medicina Interna*, **6**(2):92-97.
- CASTELAIN, S., FRANCOIS, C., BONTE, D., BARON, A., HORLE, B., MOREL, V., PAUTARD, B. & DUVERLIE, G., 2004. Epidemiological and Quantitative Study of GBV-C Infection in French Polytransfused Children. *Journal of Medical Virology*, **73**(4):596–600.
- CHANG, Q., MCLINDEN, H. J., STAPLETON, T.J., SATHAR, A. M. & XIANG, J., 2007. Expression of GB virus C NS5A protein from genotypes 1, 2, 3 and 5 and a 30 aa NS5A fragment inhibit human immunodeficiency virus type 1 replication in a CD4+ T-lymphocyte cell line. *Journal of General Virology*, **88**(12):3341–3346.

- CHARREL, N. R., MICCO, P. & LAMBALLERIE, X. 1999. Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. *Journal of General Virology*, **80**(9):2329–2335.
- CHAVEZ, D., GUERRA, B. & LANFORD, R. E., 2009. Antiviral activity and host gene induction by tamarin and marmoset interferon-alpha and interferon-gamma in the GBV-B primary hepatocyte culture model. *Virology*, **390**(2):186-96.
- CHEUNG, R. C., KEEFFE, E. B., & GREENBERG, H. B., 1997. Hepatitis G virus: is it a hepatitis virus? *Western Journal of Medicine*, **167**(1):23-33.
- CHEVALIEZ, S. & PAWLOTSKY, JM. 2006. Hepatitis C Viruses: Genomes and Molecular Biology; Chapter 1 HCV Genome and Life Cycle. Seng-Lai, editor Norfolk, UK: Horizon Bioscience;
- CHOO, Q. L., KUO, G., WEINER, A. J., OVERBY, L. R., BRADLEY, D. W. & HOUGHTON, M., 1989. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science*; **244**(4902):359-62.
- CHRISTENSEN, P. B., FISKE, N., MYGIND, L. H., KRARUP, H. B., WEDDERKOPP, N., VARMING, K. & GEORGESEN J., 2003. GB Virus C Epidemiology in Denmark: Different Routes of Transmission in Children and Low- and High-Risk Adults. *Journal of Medical Virology*, **70**(1):156–162.
- CUCEANU, M. N., TUPLIN, A. & SIMMONDS. P. 2001. Evolutionarily conserved RNA secondary structures in coding and non-coding sequences at the 3' end of the hepatitis G virus/GB-virus C genome. *Journal of General Virology*, **82**(4):713–722.
- DAVISON, S. M., SKIDMORE, S. J., COLLINGHAM, K. E., IRVING, W. L., HUBSCHER, S. G. & KELLY, D. A., 1998. Chronic hepatitis in children after liver transplantation: role of hepatitis C virus and hepatitis G virus infections. *Journal of Hepatology*, **28**(5):764–70.
- DAWSON, G. J., SCHLAUDER, G. G., PILOT-MATIAS, T. J., THIELE, D., LEARY, T. P., MURPHY, P., ROSENBLATT, J. E., SIMONS, J. N., MARTINSON, F. E., GUTIERREZ, R. A., LENTINO, J. R., PACHUCKI, C., MUERHOFF, A. S., WIDELL, A., TEGTMEIER, G., DESAI, S. & MUSHAHWAR, I. K., 1996. Prevalence studies of GB virus-C using reverse-transcriptase-polymerase chain reaction. *Journal of Medical Virology*, **50**(1):97-103.
- DEINHARDT, F., HOLMES, A. W., CAPPS, R. B. & POPPER, H., 1967. Studies on the transmission of human viral hepatitis to marmoset monkeys. I. Transmission of disease, serial passages, and description of liver lesions. *Journal of Experimental Medicine*, **125**(4): 673-88.
- DEINHARDT, F., PETERSON, D., CROSS, G., WOLFE, L. & HOLMES, A. W., 1975. Hepatitis in marmosets. *American Journal of the Medical Sciences*, **270**(1):73-80.

- DEKA, N., SHARMA, D. M. & MUKERJEE, R., 1994. Isolation of the Novel Agent from Human Stool Samples That Is Associated with Sporadic Non-A, Non-B Hepatitis. *Journal of Virology*, **68**(12):7810-7815.
- DESASSIS, J. F., LAPERCHE, S., GIRAULT, A., KOLKO, A., BOUCHARDEAU, F., ZINS, B., POIGNET, J. L & COUROUCÉ, A. M., 1999. Prevalence of present and past hepatitis G virus infection in a French haemodialysis centre. *Nephrology Dialysis Transplantation*, **14**(11):2692-7.
- DEVEREUX, H., SABIN, C. A., KINSON, Z., BROWN, D., GRIFFIOEN, A., DUSHEIKO, G. M. & LEE, C. A. 1998 Influence of HIV-1 infection on GBV-C infection in multiply infected haemophilic patients. *Journal of Medical Virology*, **56**(4):316-320.
- EL-ZAYADI, A. R., ABE, K., SELIM, O., NAITO, H., HESS, G. & AHDY, A., 1999. Prevalence of GBV-C/hepatitis G virus viraemia among blood donors, health care personnel, chronic non-B non-C hepatitis, chronic hepatitis C and haemodialysis patients in Egypt. *Journal of Virological Methods*, **80**(1):53-8.
- ERKER. C., SIMONS, J., MUERHOFF, A., LEARY, T., CHALMERS, M., DESAI, S. & MUSHAHWAR, I., 1996. Molecular cloning and characterization of a GB virus C isolate from a patient with non-A-E hepatitis. *Journal of General Virology*, **77**(11):2713-2720.
- ERKER^a, J. C., DESAI, S. M., LEARY, T. P., CHALMERS, M. L., MONTES, C. C. & EMUSHAHWAR, I. K., 1998. Genomic analysis of two GB virus A variants isolated from captive monkeys. *Journal of General Virology*, **79**(1):41-45.
- ERKER^b, C. J., DESAI, M. S. & MUSHAHWAR, K. I., 1998. Rapid detection of GB virus C RNA by reverse transcription-polymerase chain reaction (RT-PCR) using primers derived from the 5' nontranslated region. *Journal of Virological Methods*, **70**(1):1-5.
- ESTEVEES, A., PARREIRA, R., PIEDADE, J., VENENNO, T., FRANCO, M., GERMANO DE SOUSA, J., PATRÍCIO, L., BRUM, P., COSTA, A. & CANAS-FERREIRA, W. F., Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Research and Human Retroviruses*, **19**(6):511-517.
- ESTEVEES, A., PARREIRA, R., VENENNO, T., FRANCO, M., PIEDADE, J., GERMANO DE SOUSA, J. & CANAS-FERREIRA, W. F., 2002. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Research and Human Retroviruses*, **18**(5):313-25.
- FEINSTONE, S. M., KAPIKIAN, A. Z., PURCELL, R. H., ALTER, H. J. & HOLLAND, P. V., 1975. Transfusion-associated hepatitis not due to viral hepatitis type A or B. *New England Journal of Medicine*, **292**(15):767-770.

- FEINSTONE, S. M., KAPIKIAN, A. Z. & PURCELL, R. H., 1973. Hepatitis A: detection of a virus-like antigen associated with acute illness. *Science*, **182**(116):1026-1028.
- FELSENSTEIN, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, **39**(4):783-791.
- FERRAZ, M. L., SILVA, A. E., MACDONALD, G. A., TSAREV, S. A., DI BISCEGLIE, A. M. & LUCEY, M. R., 1966. Fulminant hepatitis in patients undergoing liver transplantation: evidence for a non-A, non-B, non-C, non-D, and non-E syndrome. *Liver Transplantation and Surgery*, **2**(1):60-66.
- FEUCHT, H. H., ZÖLLNER, B., POLYWKA, S., KNÖDLER, B., SCHRÖTER, M., NOLTE, H. & LAUFS, R., 1997. Prevalence of hepatitis G viremia among healthy subjects, individuals with liver disease, and persons at risk for parenteral transmission. *Journal of Clinical Microbiology*, **35**(3):767-768.
- FREITAS, J., 2003. Hepatites víricas perspectiva histórica. Disponível em <http://www.AIDSportugal.com>.
- FREY, S. E., HOMAN, S. M., SOKOL-ANDERSON, M., CAYCO, M. T., CORTORREAL, P., MUSIAL, C. E & DI BISCEGLIE, A., 2002. Evidence for probable sexual transmission of the hepatitis g virus. *Clinical Infectious Diseases*, **34**(8):1033-8.
- FUJISAWA, T., HORIIKE, N., MICHITAKA, N. & ONJI, M., 2000. HEPATITIS G VIRUS: Influence of RNA titre and amino acid changes in the NS5A region of GB virus C/hepatitis G virus on the effectiveness of interferon therapy. *Journal of Gastroenterology and Hepatology* **15**(6):632-639.
- GASCUEL, O. & STEEL, M. 2006. Neighbor-Joining Revealed. Review. *Molecular Biology and Evolution*. **23**(11):1997-2000.
- GEORGE, L. S. & VARMAZ, D. 2005. What You Need to Know About GB Virus C. *Current Gastroenterology Reports*, **7**(1):54-62.
- GEORGE, L. S., XIANG, J. & STAPLETON, T. J., 2003. Clinical isolates of GB virus type C vary in their ability to persist and replicate in peripheral blood mononuclear cell cultures. *Virology*, **316**(2):191-201.
- GHIBAUDO, D., COHEN, L., PENIN, F. & MARTIN, A., 2004. Characterization of GB Virus B Polyprotein Processing Reveals the Existence of a Novel 13-kDa Protein with Partial Homology to Hepatitis C Virus p7 Protein. *Journal of Biological Chemistry*, **279**(24):24965-24975.

- GIULIVI, A., SLINGER, R., TEPPER, M., SHER, G., SCALIA, V., KESSLER, G. & GILL, P., 2000. Prevalence of GBV-C/Hepatitis G Virus Viremia and Anti-E2 in Canadian Blood Donors. *Vox Sanguinis*, **79**(4):201–205.
- GUTIERREZ, R. A., DAWSON, G. J., KNIGGE, M. F., MELVIN, S. L., HEYNEN, C. A., KYRK, C. R., YOUNG, C. E., CARRICK, R. J., SCHLAUDER, G. G., SUROWY, T. K., DILLE, B. J., COLEMAN, P. F., THIELE, D. L., LENTINO, J. R., PACHUCKI, C. & MUSHAHWAR, I. K., 1997. Seroprevalence of GB virus C and persistence of RNA and antibody. *Journal of Medical Virology*, **53**(2):167-173.
- HALL, B. G., 2005. Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Molecular Biology and Evolution*, **22**(3):792-802.
- HALL, T. A., 1999. Bioedit, a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**:95-98.
- HANDA, A. & BROWN, K. E., 2000. GB virus C/hepatitis G virus replicates in human haematopoietic cells and vascular endothelial cells. *Journal of General Virology*, **81**(10):2461-2469.
- HANDELSMAN, E., CHENG, I., THOMPSON, B., HERSHOW, R., MOFENSON, L. M., HOLLINGER, F. B., CHEN, K.T., BURCHETT, K. D. & STAPLETON, J. T., 2007. Impact of GB virus type C infection on mother-to-child HIV transmission in the Women and Infants Transmission Study Cohort. *HIV Medicine*. **8**(8):561-7.
- HATTORI, J., OKUMURA, N., YAMAZAKI, Y., UCHIYAMA, M., HAMAGUCHI, M., NISHIYAMA, Y. & KANEDA, T., 2007. Beneficial effect of GB-virus C Co-Infection in Human Immunodeficiency Virus Type 1-Infected individuals. *Microbial Immunology*, **51**(2):193-200.
- HATTORI, J., IBE, S., NAGAI, H., WADA, K., MORISHITA, T., SATO, K., UTSUMI, M. & KANEDA, T., 2003. Prevalence of Infection and Genotypes of GBV-C/HGV among Homosexual Men. *Microbial immunology*, **47**(10):759-763.
- HAYASHI, J., FURUSYO, N., SAWAYAMA, Y., KISHIHARA, Y., KAWAKAMI, Y., ARIYAMA, I., ETOH, Y. & KASHIWAGI, S., 1998. Hepatitis G virus in general population and patients on hemodialysis. *Digestive Diseases Sciences*, **43**(9): 2143-2148.
- HEKMAT, S., MOHRAZ, M., VAHABPOUR, R., JAM, S., BAHRAMALI, G., BANIFAZL, M., AGHAKHANI, A., ESLAMIFAR, A., MAHBOUDI, F., EDALAT, R. & RAMEZANI, A., 2008. Frequency and Genotype of GB Virus C Among Iranian Patients Infected With HIV. *Journal of Medical Virology*, **80**(11):1941-1946.
- HERINGLAKE, S., TILLMANN, H. L., CORDES-TEMME, P., TRAUTWEIN, C., HUNSMANN, G. & MANNS, M. P., 1996. GBV-C/HGV is not the major cause of autoimmune hepatitis. *Journal of Hepatology*; **25**(6):980-984.

- HERRERA, E., GOMARA, J. M., MAZZINI, S., RAGG, E. & HARO, I. 2009. Synthetic Peptides of Hepatitis G Virus (GBV-C/HGV) in the Selection of Putative Peptide Inhibitors of the HIV-1 Fusion Peptide. *Journal of Physical Chemistry*, **113**(20):7383–7391.
- HIGGS, G. P. & ATTWOOD, K. T. 2005. *Bioinformatics and molecular evolution*. Blackwell publishing.
- HOFFMANN, C., ROCKSTROH, K. J. & KAMPS, S. B., 2007. *HIV Medicine*. 15th Edition, Flying Publisher. Disponível em <http://www.HIVMedicine.com>.
- HOSEINI, M., POURFATHOLLAH, A.A., MOHRAZ, M., SOHEILI, Z., AMINI, S., AGHAIEPOUR, M., SAMIEE, S., NIKOOGOFAR, M. & MESHKANI, R., 2007. Evaluation of Circulating Natural Type 1 Interferon-producing Cells in HIV/GBV-C and HIV/HCV Coinfected Patients: A Preliminary Study. *Archives of Medical Research*, **38**(8) 868-875.
- HOSSEINI-MOGHADDAM, S. M., KEYVANI, H., SAMADI, M., ALAVIAN, S. M., MAHDAVIMAZDEH, M., DANESHVAR, S. & RAZZAGHI, Z., 2008. GB Virus Type C Infection in Hemodialysis Patients Considering Co-Infection with Hepatitis C Virus. *Journal of Medical Virology*, **80**(7):1260–1263.
- HOWARD, C. R., 2002. Hepatitis viruses: a Pandora`s box? *Journal of Gastroenterology and Hepatology*, **17** Suppl:S464-S467.
- HUELSENBECK, J. P. & RONQUIST, F., 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**(8):754-755.
- IBÁÑEZ, A., GIMÉNEZ-BARCONS, M., TAJAHUERCE, A., TURAL, C., SIRERA, G., CLOTET, B., SÁNCHEZ-TAPIAS, J. M., RODÉS, J., MARTÍNEZ, M. A. & SAIZ, J. C., 1998. Prevalence and Genotypes of GB Virus C/Hepatitis G Virus (GBV-C/HGV) and Hepatitis C Virus Among Patients Infected With Human Immunodeficiency Virus: Evidence of GBV-C/HGV Sexual Transmission. *Journal of Medical Virology*, **55**(4):293–299.
- JARVIS, L. M., DAVIDSON, F., HANLEY, J. P., YAP, P. L., LUDLAM, C. A. & SIMMONDS, P., 1996. Infection with hepatitis G virus among recipients of plasma products. *Lancet*, **348**(9038):1352-1355.
- JONES, F. J., KULKARNI, S. P., BUTERA, T. S. & REEVES, C. W., 2005. GB virusC: a virus without a disease: We cannot give it chronic fatigue syndrome. *BMC Infectious Diseases*, **5**:78.
- JUNG, S., EICHENMÜLLER, M., DONHAUSER, N., NEIPEL, F., ENGEL, A. M., HESS, G., FLECKENSTEIN, B. & REIL, H., 2007. HIV entry inhibition by the envelope 2 glycoprotein of GB virus C. *AIDS*, **21**(5):645-647.

- JUNG, S., KNAUER, O., DONHAUSER, N., EICHENMÜLLER, M., HELM, M., FLECKENSTEIN, B & REIL, H., 2005. Inhibition of HIV strains by GB virus C in cell culture can be mediated by CD4 and CD8 T-lymphocyte derived soluble factors. *AIDS*, **19**(12):1267–1272.
- KAO, J. H., CHEN, P. J., LAI, M. Y., CHEN, W. & CHEN, D. S., 1998. Effect of GB virus-C/hepatitis G virus on hepatitis B and C viremia in multiple hepatitis virus infections. *Archives of Virology*, **143**(4):787–802.
- KAO, J. H., CHEN, W., CHEN, P. J., LAI, M. Y. & CHEN, D. S., 1999. Liver and peripheral blood mononuclear cells are not major sites for GB virus-C/hepatitis G virus replication. *Archives of Virology*, **144**(11):2173-2183.
- KAO, J. H., LIU, C. J., CHEN, P. J., CHEN, W., HSIANG, S. C., LAI, M. Y. & CHEN, D. S. 1997. Interspousal transmission of GB virus C/hepatitis G: a comparison with hepatitis C virus. *Journal of Medical Virology*, **53**(4):348–353.
- KARAYIANNI, P. & THOMAS, C. H. 1997. Current Status of Hepatitis G Virus (GBV-C) in Transfusion: Is It Relevant? *Vox Sanguinis*, **73**(2):63–69.
- KATAYAMA, K., KAGEYAMA, T., FUKUSHI, S., HOSHINO, F. B., KURIHARA, C., ISHIYAMA, N., OKAMURA. H. & OYA, A., 1998. Full-length GBV-C/HGV genomes from nine Japanese isolates: characterization by comparative analyses. *Archives of Virology* **143**(6):1063-1075.
- KATOH, K. & TOH, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**(4):286-298.
- KAUFMAN, T. M., MCLINDEN, J. H., XIANG, J., ENGEL, A. M. & STAPLETON, J. T., 2007. The GBV-C envelope glycoprotein E2 does not interact specifically with CD81. *AIDS*, **21**(8):1045-1048.
- KAYE, S., HOWARD, M., ALABI, A., HANSMANN, A., WHITTLE, H., & VAN DER LOEF, S. M., 2005. No Observed Effect of GB Virus C Coinfection on Disease Progression in a Cohort of African Woman Infected with HIV-1 or HIV-2. *Clinical Infectious Diseases*, **40**(6):876–878.
- KELLY, D & SKIDMORE, S., 2002. Hepatitis C-Z: recent advances. *Archives of Disease in Childhood*, **86**(5):339–343.
- KIMURA, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**(2):111-120.

- KONOMI, N., MIYOSHI, C., ZERAIN, C., LI, T. C., ARAKAWA, Y. & ABEL, K., 1999. Epidemiology of Hepatitis B, C, E, and G Virus Infections and Molecular Analysis of Hepatitis G Virus Isolates in Bolivia. *Journal of Clinical Microbiology*, **37**(10):3291–3295.
- KOPETZKI, E., JEKLE, A., JI, C., RAO, E., ZHANG, J., FISCHER, S., CAMMACK, N., SANKURATRI, S., & HEILEK, G., 2008. Closing two doors of viral entry: Intramolecular combination of a coreceptor- and fusion inhibitor of HIV-1. *Virology Journal*, **5**:56.
- KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., FOROOTAN, A., JONÁK, J., LIND, K., SINDELKA, R., SJÖBACK, R., SJÖGREEN, B., STRÖMBOM, L., STÅHLBERG, A. & ZORIC, N., 2006. The real-time polymerase chain reaction. Review. *Molecular Aspects of Medicine*, **27**(2-3):95-125.
- KUBO, S., NISHIGUCHI, S., KUROKI, T., HIROHASHI, K., TANAKA, H., TSUKAMOTO, T., SHUTO, T. & KINOSHITA, H., 1997. Poor association of GBV-C viremia with hepatocellular carcinoma. *Journal of Hepatology*, **27**(1):91-95.
- LASKUS, T., RADKOWSKI, M., WANG, L. F., VARGAS, H. & RAKELA, J., 1998. Detection of hepatitis G virus replication sites by using highly strand-specific Tth-based reverse transcriptase PCR. *Journal of Virology*, **72**(4):3072-3075.
- LASKUS, T., RADKOWSKI, M., WANG, L. F., VARGAS, H. & RAKELA, J., 1997. Lack of evidence for hepatitis G virus in the livers of patients co-infected with hepatitis C and G viruses. *Journal of Virology*, **71**(10):7804–6.
- LEARY, P. T., MUERHOFF, S. A., SIMONS, N. J., PILOT-MATIAS, J. T., ERKER, C. J., CHALMERS, L. M., SCHLAUDER, G. G., DAWSON, J. G., DESAI, M. S., & MUSHAWAR, K. I. 1996. Sequence and Genomic Organization of GBV-C: A Novel Member of the Flaviviridae Associated With Human Non-A-E. *Journal of Medical Virology*, **48**(1):60-67.
- LEFRÈRE^a, J. J., ROUDOT-THORAVALE, F., MORAND-JOUBERT, L., BROSSARD, Y., PARNET-MATHIEU, F., MARIOTTI, M., AGIS, F., ROUET, G., LERABLE, J., LEFÈVRE, G., GIROT, R. & LOISEAU, P., 1999. Prevalence of GB virus type C/hepatitis G virus RNA and of anti-E2 in individuals at high or low risk for blood-borne or sexually transmitted viruses: evidence of sexual and parenteral transmission. *Transfusion*. **39**(1):83-94.
- LEFRÈRE^b, J. J., FÉREC, C., ROUDOT-THORAVALE, F., LOISEAU, P., CANTALOUBE, J. F., BIAGINI, P., MARIOTTI, M., LEGAC, G. & MERCIER, B., 1999. GBV-C/Hepatitis G Virus (HGV) RNA Load in Immunodeficient Individuals and in Immunocompetent Individual. *Journal of Medical Virology*, **59**(1):32–37.
- LEFRÈRE, J. J., SENDER, A., MERCIER, B., MARIOTTI, M., PERNOT, F., SOULIÉ, J. C., MALVOISIN, A., BERRY, M., GABAI, A., LATTES, F., GALIAY, J. C., PAWLAK, C., DE LACHAUX, V., CHAUVEAU, V., HREICHE, G., LARSEN, M., FÉREC, C., PARNET-MATHIEU, F., ROUDOT-THORAVALE, F., BROSSARD, Y., 2000. High rate of GB virus type

C/HGV transmission from mother to infant: possible implications for the prevalence of infection in blood donors. *Transfusion*, **40**(5):602-607.

LI, W. H., WU, C. I. & LUO, C. C., 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution*, **2**(2):150-174.

LI^a, C., COLLINI, P., DANSO, K., OWUSU-OFORI, S., DOMPREH, A., CANDOTTI, D., OPARE-SEM, O. & ALLAIN, J. P., 2006. GB virus C and HIV-1 RNA load in single virus and co-infected West African individuals. *AIDS*, **20**(3):379–386.

LI^b, C., DANSO, K., ADDO-YOBO, E., DOMPREH, A., SARKODIE, F., OWUSU-OFORI, S., ALLAIN, J. P., 2006. GB virus C genotype 1 is rarely transmitted vertically but acquired during infancy in West Africa. *AIDS*, **20**(10):1458-1460.

LIMA, V. D., FINK, V., YIP, B., HOGG, R. S., HARRIGAN, P. R. & MONTANER, J. S., 2009. Association between HIV-1 RNA level and CD4 cell count among untreated HIV-infected individuals. *American Journal of Public Health*, **99** Suppl 1:S193-6.

LINDENBACH, B. D., RICE, C. M. & THIEL, H. J., 2007. *Flaviviridae: The viruses and Their replication*. Fields Virology, D. M. Knipe and P. M. Howley, 5th Edition Lippincott-Raven Publishers, Philadelphia, 1103-1125 pp.

LINNEN, J., WAGES, J. JR., ZHANG-KECK, Z. Y., FRY, K. E., KRAWCZYNSKI, K. Z., ALTER, H., KOONIN, E., GALLAGHER, M., ALTER, M., HADZIYANNIS, S., KARAYIANNIS, P., FUNG, K., NAKATSUJI, Y., SHIH, J. W., YOUNG, L., PIATAK, M. JR., HOOVER, C., FERNANDEZ, J., CHEN, S., ZOU, J. C., MORRIS, T., HYAMS, K. C., ISMAY, S., LIFSON, J. D., HESS, G., FOUNG, S. K., THOMAS, H., BRADLEY, D., MARGOLIS, H., KIM, J. P., 1996. Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science*. **271**(5248):505-8.

LIU, H. F., TENG, C. W., FUKUDA, Y., NAKANO, I., HAYASHI, K., TAKAMATSU, J., GOUBAU, P. & TOYODA, H., 2003. A novel subtype of GB virus C/hepatitis G virus genotype 1 detected uniquely in patients with hemophilia in Japan. *Journal of Medical Virology*, **71**(3):385–90.

LIU, H. F., MUYEMBE-TAMFUM, J. J., DAHAN, K., DESMYTER, J. E GOUBAU, P. 2000. High prevalence of GB virus C/hepatitis G virus in Kinshasa, Democratic Republic of Congo: a phylogenetic analysis. *Journal of Medical Virology*, **60**(2):159-165.

LOLE, K. S., BOLLINGER, R. C., PARANJAPPE, R. S., GADKARI, D., KULKARNI, S. S., NOVAK, N. G., INGERSOLL, R., SHEPPARD, H. W. & RAY, S. C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology*, **73**(1):152-160.

- LOPEZ-ALCOROCHO, M. J., FOGEDA, M., CASTILLO, I., BARTOLOMÉ, J. & CARREÑO, V., 1998. Phylogenetic Analysis of Hepatitis GB Virus Type C/Hepatitis G Virus in Spanish Patients with Chronic Hepatitis B or C Virus Infection. *Intervirology*, **41**(6):233–237.
- LOUREIRO, C. L., ALONSO, R., PACHECO, B. A., UZCÁTEGUI, M. G., VILLEGAS, L., LEÓN, G., DE SAÉZ, A., LIPRANDI, F., LÓPEZ, J. L., PUJOL, F. H., 2002. High prevalence of GB virus C/hepatitis G virus genotype 3 among autochthonous Venezuelan populations. *Journal of Medical Virology*, **68**(3):357-362.
- LU, L., NG, M. H., ZHOU, B., LUO, H., NAKANO, T., ROBERTSON, B. H. & IM, S. W., 2001. Detection and genotyping of GBV-C/HGV variants in China. *Virus Research*, **73**(2):131-44.
- MACKAY, M. I., ARDEN, E. K. & NITCSHE .2002. Real-time PCR in *Virology*. *Nucleic Acids Research*, **3**(6):1292-1305.
- MAIDANA-GIRET, M. T., SILVA, T. M., SAUER, M. M., TOMIYAMA, H., LEVI, J. E., BASSICHETTO, K. C., NISHIYA, A., DIAZ, R. S., SABINO, E. C., PALACIOS, R., KALLAS, E. G., 2009. GB virus type C infection modulates T-cell activation independently of HIV-1 viral load. *AIDS*, **23**(17):2277-2287.
- MARMOR, M., HERTZMARK, K., THOMAS, S. M., HALKITIS, P. N. & VOGLER, M., 2006. Resistance to HIV infection. *Journal of Urban Health*, **83**(1):5-17.
- MARTINI, F., SACCHI, A., LALLE, E., CASTILLETI, C., D'OFFIZI, G., ABBATE, I. & CAPOBIANCHI, M. R. 2008. GB Virus Type C–Driven Protection in HIV/HCV Coinfection: Possible Role of Interferon Gamma and Dendritic Cell Activation. *Gastroenterology*, **134**(5):1631-1633.
- MASUKO, K., MITSUI, T., IWANO, K., YAMAZAKI, C., OKUDA, K., MEGURO, T., MURAYAMA, N., INOUE, T., TSUDA, F., OKAMOTO, H., MIYAKAWA, Y. & MAYUMI, M., 1996. Infection with hepatitis GB virus C in patients on maintenance haemodialysis. *New England Journal of Medicine*, **334**(23):1485-1490.
- MATHET, V. L., ESPÍNOLA, L., RUIZ, V., MARÍNCOLA, A., QUARLERI, J. F., CEBALLOS, A., PERALTA, L. A., NATAL, M., HAEDO, A., SÁNCHEZ, D. O. & OUBIÑA, JR., 2003. Phylogenetic and Mathematical Analyses for Investigating Putative Mother-to-Infant Transmission Chains When Only GB Virus C (Hepatitis G Virus) 5'Noncoding Region Sequences Are Available. *Journal of Clinical Microbiology*, **41**(9):4489–4491.
- MAUSER-BUNSCHOTEN, E. P., DAMEN, M., ZAAIJER, H. L., SJERPS, M., ROOSENDAAL, G., LELIE, P. N., CUYPERS, H. T. & VAN DEN BERG, H. M., 1998. Hepatitis G virus RNA and Hepatitis G virus-E2 antibodies in Dutch hemophilia patients in relation to transfusion history. *Blood*, **92**(6):2164–8.

- MEHRISHI, N. J. & BAKÁCS, T. 2005. HIV and hepatitis G virus/GB virus C co-infection: beneficial or not? *Lancet Infectious Diseases*, **5**(8):464-465.
- MENDONZA, C. & SORIANO, V. 2009. Update on HIV Viral-load Assays: New Technologies and Testing in Resource-limited Settings Future. *Virology*, **4**(5):423-430.
- MISON, L., HYLAND, C., POIDINGER, M., BORTHWICK, I., FAOAGALI, J., AENO, U. & GOWANS, E., 2000. Hepatitis G virus genotypes in Australia, Papua New Guinea and the Solomon Islands: A possible new Pacific type identified. *Journal of Gastroenterology and Hepatology*, **15**(8):952-956.
- MOENKEMEYER, M., SCHMIDT, E. R., WEDEMEYER, H., TILLMANN, L. H. & HEIKEN, H., 2008. GBV-C Coinfection Is Negatively Correlated to Fas Expression and Fas-Mediated Apoptosis in HIV-1 Infected Patients. *Journal of Medical Virology*, **80**(11):1933-1940.
- MOHR, L. E. & STAPLETON, T. J. 2009. GB virus type C interactions with HIV: the role of envelope Glycoproteins. *Journal of Viral Hepatitis*, **16**(11):757-768.
- MORADPOUR, D., PENIN, F. & RICE, M. C., 2007. Replication of hepatitis C virus. *Nature Reviews Microbiology*, **5**(6):453-463.
- MOSAM, A., SATHAR, M. A., DAWOOD, H., CASSOL, E., ESTERHUIZEN, T. M. & COOVADIA, H. M., 2007. Effect of GB virus C co-infection on response to generic HAART in African patients with HIV-1 clade C infection. *AIDS*, **21**(10):1377-1379.
- MPHAHLELE, M. J., ASPINALL, S., SPOONER, R. & CARMAN, W. F., 1999. Age related prevalence of hepatitis G virus in South Africans. *Journal of Clinical Pathology*, **52**(10):752-757.
- MUERHOFF, A. S., SIMONS, J. N., LEARY, T. P., ERKER, J. C., CHALMERS, M. L., PILOT-MATIAS, T. J., DAWSON, G. J., DESAI, S. M. & MUSHAHWAR, I. K., 1996. Sequence heterogeneity within the 5'-terminal region of the hepatitis GB virus C genome and evidence for genotypes. *Journal of Hepatology*, **25**(3):379-384.
- MUERHOFF, A. S., TILLMANN, H. L., MANNS, M. P., DAWSON, G. J. & DESAI, S. M., 2003. GB virus C genotype determination in GB virus-C/HIV co-infected individuals. *Journal of Medical Virology*, **70**(1):141-149.
- MUERHOFF, S., DAWSON, J. G. & DESAI, M. S., 2006. A Previously Unrecognized Sixth Genotype of GB Virus C Revealed by Analysis of 5'-Untranslated Region Sequences. *Journal of Medical Virology*, **78**(1):105-111.

- MUERHOFF, S., LEARY, P. T., SATHAR, A. M., DAWSON, J. G. & DESAI M. S., 2005. African origin of GB virus C determined by phylogenetic analysis of a complete genotype 5 genome from South Africa. *Journal of General Virology*, **86**(6):1729–1735.
- MUERHOFF, A. S., LEARY, T. P., SIMONS, J.N., PILOT-MATIAS, T. J., DAWSON, G. J., ERKER, J. C., CHALMERS, M. L., SCHLAUDER, G. G., DESAI, S. M. & MUSHAHWAR, I. K., 1995. Genomic Organization of GB Viruses A and B: Two New Members of the Flaviviridae Associated with GB Agent Hepatitis. *Journal of Virology*, **69**(9):5621–30.
- MUERHOFF, S., SMITH, B. D., LEARY, P. T., ERKER, C. J. DESAI, M. S. & MUSHAHWAR, K. I., 1997. Identification of GB Virus C Variants by Phylogenetic Analysis of 5-Untranslated and Coding Region Sequences. *Journal of Virology*, **71**(9):6501–6508.
- MUÑOZ, S. J., ALTER, H. J. & NAKATSUJI, Y., 1999. The significance of hepatitis G virus in serum of patients with sporadic fulminant and sub-fulminant hepatitis of unknown etiology. *Blood*, **94**(4):1460–4.
- NAITO, H., HAYASHI, S. & ABE, K., 2000. The entire nucleotide sequence of two hepatitis G virus isolates belonging to a novel genotype: isolation in Myanmar and Vietnam. *Journal of General Virology*, **81**(1):189–194.
- NAITO, H. & ABE, K., 2001. Genotyping system of GBV-C/HGV type 1 to type 4 by the polymerase chain reaction using type-specific primers and geographical distribution of viral genotypes. *Journal of Virological Methods*, **91**(1):3–9.
- NAITO, H., WIN, M. K. & ABE, K., 1999. Identification of a Novel Genotype of Hepatitis G Virus in Southeast Asia. *Journal of Clinical Microbiology*, **37**(4):1217–1220.
- NATTERMANN, J., NISCHALKE, H. D., KUPFER, B., ROCKSTROH, J., HESS, L., SAUERBRUCH, T. & SPENGLER, U., 2003. Regulation of CC chemokine receptor 5 in hepatitis G virus infection. *AIDS*, **17**(10):1457–1462.
- NEI, M. & GOJOBORI, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, **3**(5):418–26.
- NERURKAR, V. R., CHUA, P. K., HOFFMANN, P. R., DASHWOOD, W. M., SHIKUMA, C. M. & YANAGIHARA, R., 1998. High Prevalence of GB Virus C/Hepatitis G Virus Infection Among Homosexual Men Infected With Human Immunodeficiency Virus Type 1: Evidence for Sexual Transmission. *Journal of Medical Virology*, **56**(2):123–127.
- NUNNARI, G., NIGRO, L., PALERMO, F., ATTANASIO, M., BERGER, A., DOERR, H. W., POMERANTZ, R. J. & CACOPARDO, B., 2003. Slower progression of HIV-1 infection in persons with GB virus C co-infection correlates with an intact T-helper 1 cytokine profile. *Annals of Internal Medicine*, **139**(1):26–30.

- NYLANDER, J. A. A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- ODEH, R., AL-MOSLIHET, I. M., AL-JOKHDAR, M. & EZZEDDINE, A. S. 2005. Detection and Genotyping of GBV-C Virus in the United Arab Emirates. *Journal of Medical Virology* **76**(4):534–540.
- OHTO, H., UJIIE, N., SATO, A., OKAMOTO, H. & MAYUMI, M., 2000. Mother to-infant transmission of GB virus type C/HGV. *Transfusion*, **40**(6):725-730.
- OKAMOTO, H., NAKAO, H., INOUE, T., FUKUDA, M., KISHIMOTO, J., IIZUKA, H., TSUDA, F., MIYAKAWA, Y. & MAYUMI, M., 1997. The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. *Journal of General Virology*, **78**(4):737–745.
- ORII, K., TANAKA, E., ROKUHARA, A., MARUYAMA, A., ICHIJO, T., YOSHIZAWA, K. & KIYOSAWA, K., 2000. Persistent infection mechanism of GB virus C/hepatitis G virus differs from that of hepatitis C virus. *Intervirology*, **43**(3):139–145.
- ORITO, E., MIZOKAMI, M., YASUDA, K., SUGIHARA, K., NAKAMURA, M., MUKAIDE, M., OHBA, K. I., NAKANO, T., KATO, T., KONDO, Y., KUMADA, T., UEDA, R. & IINO, S., 1997. Interferon-alpha therapy in patients dually infected with hepatitis C virus and GB virus C/hepatitis G virus: Virological response of HGV and pretreatment HGV viremia level. *Journal of Hepatology*, **27**(4): 603–12.
- PARREIRA, R., PÁDUA, E., PIEDADE, J., VENENNO, T., PAIXÃO, M. T. & ESTEVES, A., 2005. Genetic analysis of human immunodeficiency virus type 1 nef in Portugal: subtyping, identification of mosaic genes, and amino acid sequence variability. *Journal of Medical Virology*, **77**(1):8-16.
- PATERNOSTER, D., SERENA, A., SANTIN, M., MARCHIORI, S., AMORUSO, E., LONGO, D. & GUSSETTI, N., 2009. GB virus C infection in pregnancy: Maternal and perinatal importance of the infection. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, **144**(2):115-118.
- PAVESI, A. 2001. Origin and Evolution of GBV-C/Hepatitis G Virus and Relationships with Ancient Human Migrations. *Journal of Molecular Evolution*, **53**(2):104–113.
- PESSOA, M. G., TERRAULT, N. A., DETMER, J., KOLBERG, J., COLLINS, M., HASSOBA, H. M. & WRIGHT, T. L., 1998. Quantitation of hepatitis G and C viruses in the liver: evidence that hepatitis G virus is not hepatotropic. *Hepatology*, **7**(3):877-880.
- PETERLIN, M. B. & TRONO, D., 2003. Hide, shield and strike back: how HIV-infected cells avoid immune eradication. *Nature Reviews. Immunology*. **3**(2):97-107.

- PETRIK, J., GUELLA, L., WIGHT, D. G. D., PEARSON, M. G., HINTON, J., PARKER, H., ALLAIN, J. P. & ALEXANDER, J. M. G. 1998. Hepatic histology in hepatitis C virus carriers coinfecting with hepatitis G virus. *Gut*, **42**(1):103–106.
- PILOT-MATIAS, T. J., CARRICK, R. J., COLEMAN, P. F., LEARY, T. P., SUROWY, T. K., SIMONS, J. N., MUEHROFF, A. S., BUIJK, S. L., CHALMERS, M. L., DAWSON, G. J., DESAI, S. M. & MUSHAHWAR, I. K., 1996. Expression of the GB Virus C E2 Glycoprotein Using the Semliki Forest Virus Vector System and Its Utility as a Serologic Marker. *Virology*, **225**(2):282–292.
- PINHO, J. R., ZANOTTO, P. M., FERREIRA, J. L., SUMITA, L. M., CARRILHO, F. J., DA SILVA, L. C., CAPACCI, M. L., SILVA, A. O., GUZ, B., GONÇALES, F. L. JR, GONÇALES, N. S., BUCK, G. A., MEYERS, G. A. & BERNARDINI, A. P., 1999. High Prevalence of GB Virus C in Brazil and Molecular Evidence for Intrafamilial Transmission. *Journal of Clinical Microbiology*, **37**(5):1634–1637.
- PIROTH, L., CARRAT, F., LARRAT, S., GODEREL, I., MARTHA, B., PAYAN, C., LUNEL-FABIANI, F., BANI-SADR, F., PERRONNE, C., CACOU, P., POL, S. & MORAND, P., 2008. Prevalence and impact of GBV-C, SEN-V and HBV occult infections in HIV–HCV co-infected patients on HCV therapy. *Journal of Hepatology*, **49**(6):892–898.
- POLGREEN, M. P., XIANG, J., CHANG, Q. & STAPLETON, T. J., 2003. GB virus type C/hepatitis G virus: a non-pathogenic flavivirus associated with prolonged survival in HIV-infected individuals. *Microbes and Infection*, **5**(13):1255–1261.
- POLISH, L. B., GALLAGHER, M., FIELDS, H. A. & HADLER, S. C., 1993. Delta hepatitis: molecular biology and clinical and epidemiologic features. *Microbiology Clinical Microbiology Reviews*, **6**(3):211–229.
- POSADA, D. & CRANDALL, K. A., 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**(9):817–818.
- QUIROS-ROLDAN, E., QUIROS-ROLDAN, A. & MAROTO, C., 2001. Investigation of saliva, faeces, urine or semen samples for the presence of GBV-C RNA. *European journal of Epidemiology*, **17**(39):271–274.
- QURISHI, N., KREUZBERG, C., LÜCHTERS, G., EFFENBERGER, W., KUPFER, B., SAUERBRUCH, T., ROCKSTROH, J. K. & SPENGLER, U., 2003. Effect of antiretroviral therapy on liver-related mortality in patients with HIV and hepatitis C virus coinfection. *Lancet*, **362**(9397):1708–1713.
- RADKOWSKI, M., WANG, L. F., CIANCIARA, J., RAKELA, J. & LASKUS, T., 1999. Analysis of hepatitis G virus/GB virus C quasispecies and replication sites in human subjects. *Biochemical and Biophysical Research Communications*, **258**(2):296–299.

- RADKOWSKI, M., KUBICKA, J., KISIEL, E., CIANCIARA, J., NOWICKI, M., RAKELA, J. & TOMASZ LASKUS, T., 2000. Detection of active hepatitis C virus and hepatitis G virus/GB virus C replication in bone marrow in human subjects. *Blood*, **95**(12):3986-3989.
- RAMEZANI, A., MOHRAZ, M., VAHABPOUR, R., JAM, S., BANIFAZL, M., ESLAMIFAR, A., MAHBOUDI, F., AGHAKHANI, A., EDALAT, R. & HEKMAT, S., 2008. Frequency of Hepatitis G Virus Infection among HIV Positive Subjects with Parenteral and Sexual Exposure. *Journal of Gastrointestinal and Liver Diseases*, **17**(3):269-72.
- RELETFORD J. H., 2008. Genetic evidence and the modern human origins debate. *Heredity*, **100**(6):555-563.
- RENDINA, D., VIGORITA, E., BONAVOLTA, R., D'ONOFRIO, M., IURA, A., PIETRONIGRO, M. T., LACCETTI, R., BONADIES, G., LIUZZI, G., BORGIA, G., FORMISANO, P., LACCETTI, P. & PORTELLA, G., 2001. HCV and GBV-C/HGV infection in HIV positive patients in southern Italy. *European Journal of Epidemiology*, **17**(9):801-807.
- RESHETNYAK, I. V., KARLOVICH, I. T. & ILCHENKO, U. L., 2008. Hepatitis G virus. *World Journal of Gastroenterology*, **14**(30):4725-4734.
- REY, D., VIDINIC-MOULARDE, J., MEYER, P., SCHMITT, C., FRITSCH, S., LANG, J. M. & STOLL-KELLER, F., 2000. High Prevalence of GB Virus C/Hepatitis G Virus RNA and Antibodies in Patients Infected with Human Immunodeficiency Virus Type 1. *European Journal of Clinical Microbiology & Infectious Diseases*, **19**(9):721-724.
- RIZZETTO, M., CANESE, M. G., ARICÒ, S., CRIVELLI, O., TREPO, C., BONINO, F. & VERME, G., 1977. Immunofluorescence detection of a new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HbsAg carriers. *Gut*, **12**:997-1003.
- RODNEY, M., RATCLIFF, M. R., CHANG, G., TUCKWENG, K. & SLOOTS, P. T., 2007. Molecular Diagnosis of Medical Viruses. *Current Issues in Molecular Biology*, **9**:87-102.
- RODRIGUEZ, B., WOOLLEY, I., LEDERMAN, M. M., ZDUNEK, D., HESS, G. & VALDEZ, H., 2003. Effect of GB virus C coinfection on response to antiretroviral treatment in human immunodeficiency virus-infected patients. *Journal of Infectious Diseases*, **187**(3):504-507.
- RUIZ, V., ESPÍNOLA, L., MATHET, L. V., PERANDONES, E. C. & OUBIÑA, R. J. 2006. Design, development and evaluation of a competitive RT-PCR for quantitation of GBV-C RNA. *Journal of Virological Methods*, **136**(1-2):58-64.
- RUIZ, V., ESPÍNOLA, L., MATHET, V. L., PERANDONES, C. E. & OUBIÑA, J. R., 1997. Plus- and minus-stranded hepatitis G virus RNA in liver tissue and in peripheral blood mononuclear cells. *Biochemical and Biophysical Research Communications*, **237**(2):288-291.

- SAITOU, N. & NEI, M., 1987. The neighbor-joining method: a new method for reconstruction of phylogenetic trees. *Molecular Biology and Evolution*, **4**(4):406–425.
- SAMBROOK, J., FRITSCH, E. F., & MANIATIS, T., 1989. *Molecular cloning: a laboratory manual*. 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- SANGER, F. NICKLEN, S. & COULSON, R. A., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, **74**(12):5463-7.
- SATHAR, A. M. & YORK, F. D., 2001. Group5: GBV-C/HGV Isolates From South Africa. *Journal of Medical Virology*, **65**(1):121-122.
- SATHAR, A. M., SONI, N. P. & YORK, D., 2000. GB Virus C/Hepatitis G Virus (GBV-C/HGV): still looking for a disease. Review. *International Journal of Experimental Pathology*. **81**(5):305-322.
- SATHAR, M. A., SONI, P. N., PEGORARO, R., SIMMONDS, P., SMITH, D. B., DHILLON, A. P. & DUSHEIKO, G. M., 1999. A new variant of GB virus C:hepatitis G virus (GBV-C:HGV) from South Africa. *Virus Research*, **64**(2):151–160.
- SATHAR, A. M., YORK, F. D., GOUWS, E., COUTSOUDIS, A. & COOVADIA, M. H., 2004. GB Virus Type C Coinfection in HIV-Infected African Mothers and Their Infants, KwaZulu Natal, South Africa. *Clinical Infectious Diseases*, **38**(3):405-409.
- SAWAYAMA, Y., HAYASHI, J., ETOH, Y., URABE, H., MINAMI, K. & KASHIWAGI, S., 1999. Heterosexual transmission of the GB virus C/Hepatitis G virus infection to non-intravenous drug-using female prostitutes in Fukuoka, Japan. *Digestive Diseases Sciences*, **44**(10):1937–1943.
- SCALLAN, M. F., CLUTTERBUCK, D., JARVIS, L. M., SCOTT, G. R. & SIMMONDS, P., 1998. Sexual transmission of GB virus C/ hepatitis G virus. *Journal of Medical Virology*, **55**(5):203-208.
- SCHALUDER, G. G., DAWSON, G. J., SIMONS, J. N., PILOT-MATIAS, T. J., GUTIERREZ, R. A., HEYNEN, C. A., KNIGGE, M. F., KURPIEWSKI, G. S., BUIJK, S. L., LEARY, T. P., MUERHOFF, A. S., DESAI, S. M. & MUSHAHWAR, K. I., 1995. Molecular and serologic analysis in the transmission of the GB hepatitis agents. *Journal of Medical Virology*, **46**(1): 81-90.
- SCHOOLEY, T. R., 2005. HIV and Hepatitis C Virus Coinfection: Bad Bedfellows. *Topics in HIV Medicine*, **13**(4):112-116.
- SCHWARZE-ZANDER, C., BLACKARD, J. T., ZHENG, H., ADDO, M. M., LIN, W., ROBBINS, G. K., SHERMAN, K. E., ZDUNEK, D., HESS, G., CHUNG, R. T. & AIDS

CLINICAL TRIAL GROUP A5071 STUDY TEAM., 2006. GB Virus C (GBV-C) Infection in Hepatitis C Virus (HCV)/HIV–Coinfected Patients Receiving HCV Treatment: Importance of the GBV-C Genotype. *Journal of Infectious Diseases*, **194**(4):410-419.

SEIPP, S., SCHEIDEL, M., HOFMANN, W. J., TÖX, U., THEILMANN, L., GOESER, T. & KALLINOWSKI, B., 1999. Hepatotropism of GB virus C (GBV-C): GBV-C replication in human hepatocytes and cells of human hepatoma cell lines. *Journal of Hepatology*, **30**(4):570-579.

SHANKAR, E. M., SOLOMON, S. S., VIGNESH, R., MURUGAVEL, K. G., SUNDARAM, M., SOLOMON, S., BALAKRISHNAN, P. & KUMARASAMY, N., 2008. GB virus infection: a silent anti-HIV panacea within? *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **102**(12):1176-80.

SHENG, W. H., HUNG, C. C., WU, R. J., WANG, J. T., CHEN, P. J., CHANG, S. C., KAO, J. H., 2007. Clinical Impact of GB Virus C Viremia on Patients with HIV Type 1 Infection in the Era of Highly Active Antiretroviral Therapy. *Clinical Infectious Diseases*, **44**(4):584–590.

SILICIANO, F. R. 2005. *Acquired Immune Deficiency Syndrome (AIDS)*. Encyclopedia of life sciences, 1-12pp.

SIMMONDS, P. & SMITH, B. D., 1999. Structural Constraints on RNA Virus Evolution. *Journal of Virology*, **73**(7):5787-5794.

SIMMONDS^a, P. 2001. Reconstructing the origins of human hepatitis viruses. *Philosophical Transactions of the Royal Society of London Biological Sciences*, **356**(1411):1013-1026.

SIMMONDS^b P. 2001. The origin and evolution of hepatitis viruses in humans. *Journal of General Virology*, **82**(4):693-712.

SIMONS, J. N., DESAI, S. M., SCHULTZ, D. E., LEMON, S. M. & MUSHAHWAR, I. K., 1996. Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *Journal of Virology*, **70**(9):6126-6135.

SIMONS, J. N., PILOT-MATIAS, T. J., LEARY, T. P., DAWSON, G. J., DESAI, S. M., SCHLAUDER, G. G., MUERHOFF, A. S., ERKER, J. C., BUIJK, S. L., CHALMERS, M. L., VAN SANT, L. C. & MUSHAHWAR, K. I., 1995. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings of the National Academy of Sciences, USA*, **92**(8):3401-3105.

SMITH, B. D., CUCEANU, N., DAVIDSON, F., JARVIS, M. L., MOKILI, J., HAMID, S., HRISTOPHER A. LUDLAM, A. C. & SIMMONDS, P., 1997 Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5' non-coding region. *Journal of General Virology*, **78**(7):1533–1542.

- SMITH, D. B., BASARAS, M., FROST, S., HAYDON, D., CUCEANU, N., PRESCOTT, L., KAMENKA, C., MILLBAND, D., SATHAR, M. A. & SIMMONDS, P., 2000. Phylogenetic analysis of GBV-C/hepatitis G virus. *Journal of General Virology*, **81**(3):769–780.
- SORIANO-SARABIA, N., ABAD, M. A., VALLEJO, A., GUTIERREZ, S. & LEAL, M., 2006. Influence of hepatitis C and hepatitis G virus co-infection on viral and cellular dynamics in patients infected with human immunodeficiency virus following interruption of highly active anti-retroviral therapy. *Clinical Microbiology and Infection*, **12**(33):290-293.
- SOUZA, E. I., ALLEN, B. J., XIANG J., KLINZMAN, D., DIAZ, R., ZHANG, S., CHALONER, K., ZDUNEK, D., HESS, G. WILLIAMS, F.C., BENNING, L & AND J. STAPLETON J. T., 2006. Effect of Primer Selection on Estimates of GB Virus C (GBV-C) Prevalence and Response to Antiretroviral Therapy for Optimal Testing for GBV-C Viremia. *Journal of Clinical Microbiology*, **44**(9):3105–3113.
- SPELLBERG, B. & EDWARDS, J. E. JR., 2001. Type 1/type 2 immunity in infectious diseases. *Clinical Infectious Diseases*, **32**(1):76–102.
- STANYON, R., SAZZINI, M. & LUISELLI, D., 2009. Timing the first human migration into eastern Asia. Review. *Journal of Biological Chemistry*, **8**(2):18.
- STAPLETON, J. T., CHALONER, K., ZHANG, J., KLINZMAN, D., SOUZA, I. E., XIANG, J., LANDAY, A., FAHEY, J., POLLARD, R. & MITSUYASU, R., 2009. GB virus C viremia is associated with reduced CD4 expansion following interleukin 2 therapy in HIV-infected people receiving HAART. *AIDS*, **23**(5):605–610.
- STAPLETON, J. T., 2007. A New Variable Influencing HCV–Related Liver Disease in HIV–HCV Coinfected Individuals? *Gastroenterology*, **133**(6):2042-2045.
- STAPLETON, J. T., 2003. GB Virus Type C/Hepatitis G virus. *Seminars in liver disease*, **23**(2):137-148.
- STAPLETON, J. T., WILLIAMS, F. C., & XIANG, J., 2004. GB Virus Type C: a Beneficial Infection? 1. *Journal of Clinical Microbiology*, **42**(9):3915–3919.
- STARK, K., DOERING, C. D., BIENZLE, U., PAULI, G., HAMOUDA, O., ENGEL, A. M. & SCHREIER, E., 1999. Risk and clearance of GB virus C/hepatitis G virus infection in homosexual men: A longitudinal study. *Journal of Medical Virology*, **59**(3):303-306.
- STRAUSS, E., DA COSTA GAYOTTO, L. C., FAY, F., FAY, O., FERNANDES, H. S., FISCHER CHAMONE, D. A., 2002. Liver histology in co-infection of hepatitis C virus (HCV) and hepatitis G virus (HGV). *Revista do Instituto de Medicina Tropical de São Paulo*, **44**(2):67-70.

- SUZUKI, T., ISHII, K., AIZAKI, H. & WAKITA, T., 2007. Hepatitis C viral life cycle. *Advanced Drug Delivery Reviews*, **59**(12):1200–1212.
- SUZUKI, Y., KATAYAMA, K., FUKUSHI, S., KAGEYAMA, T., OYA, A., OKAMURA, H., TANAKA, Y., MIZOKAMI, M. & GOJOBORI, T., 1999. Slow Evolutionary Rate of GB Virus C/Hepatitis G Virus. *Journal of Molecular Evolution*, **8**(4):383–389.
- SWOFFORD, D. L., 2003. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0b 10. Sinauer Associates, Sunderland, Massachusetts.
- TAGGER, A., DONATO, F., RIBERO, M. L., CHIESA, R., TOMASONI, V., PORTERA, G., GELATTI, U., ALBERTINI, A., FASOLA, M. & NARDI, G., 1997. A case-control study on GB virus C/hepatitis G virus infection and hepatocellular carcinoma. Brescia HCC Study. *Hepatology*, **26**(6):1653-1657.
- TAKAHASHI, K., HIJIKATA, M., HINO, K. & MISHIRO, S., 1997. Entire polyprotein-ORF sequences of Japanese GBV-C/HGV isolates: Implications for new genotypes. *Hepatology Research*. **8**(2):139–148.
- TAMEDA, Y., KOSAKA, Y., TAGAWA, S., TAKASE, K., SAWADA, N., NAKAO, H., TSUDA, F., TANAKA, T., OKAMOTO, H., MIYAKAWA, Y. & MAYUMI, M., 1996. Infection with GB virus C (GBV-C) in patients with fulminant hepatitis. *Journal of Hepatology*, **25**(6):842-847.
- TAMURA, K., DUDLEY, J., NEI, M. & KUMAR. S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**(8):1596-1599.
- TANAKA, Y., MIZOKAMI, M., ORITO, E., OHBA, K., KATO, T., KONDO, Y., MBOUDJEKA, I., ZEKENG, L., KAPTUE, L., BIKANDOU, B., M'PELE, P., TAKEHISA, J., HAYAMI, M., SUZUKI, Y. & GOJOBORI, T., 1998. African origin of GB virus C/hepatitis G virus. *Federation of European Biochemical Societies, Letters*, **423**(2):143-148.
- TANAKA, Y., MIZOKAMI, M., ORITO, E., OHBA, K., NAKANO, T., KATO, T., KONDO, Y., DING, X., UEDA, R., SONODA, S., TAJIMA, K., MIURA, T. & HAYAMI, M., 1998. GB virus c/hepatitis g virus infection among Colombian native Indians. *American Society of Tropical Medicine and Hygiene*, **59**(3):462–467.
- THOMAS, D. L., VLAHOV, D., ALTER, H. J., HUNT, J. C., MARSHALL, R., ASTEMBORSKI, J. & NELSON, K. E., 1998. Association of antibody to GB virus C (hepatitis G virus) with viral clearance and protection from reinfection. *Journal of Infectious Diseases*; **177**(3):539-542.
- THURNER, C., WITWER, C., HOFACKER, L. I. & STADLER, F. P., 2004. Conserved RNA secondary structures in Flaviviridae genomes. *Journal of General Virology*, **85**(5)1113–1124.

- TILLMANN, H. L., MANNS, M. P., CLAES, C., HEIKEN, H., SCHMIDT, R. E. & STOLL, M., 2004. GB virus C infection and quality of life in HIVpositive patients. *AIDS*, **16**(6):736–43.
- TILLMANN, H. L., HERINGLAKE, S., TRAUTWEIN, C., MEISSNER, D., NASHAN, B., SCHLITT, H. J., KRATOCHVIL, J., HUNT, J., QIU, X., LOU, S. C., PICHLMAYR, R. & MANNS, M. P., 1998. Antibodies against the GB virus C envelope 2 protein before liver transplantation protect against GB virus C de novo infection. *Hepatology*, **28**(2):379–384.
- TILLMANN, L. H. & MANNS, P. M. 2001.GB virus-C infection in patients infected with the human immunodeficiency virus. *Antiviral Research*, **52**(2):83–90.
- TILLMANN, L. H., PATEL, K. & MCHUTCHISON G. J., 2008. GBV-C in HIV-HCV Coinfection, Low Liver-Related Mortality But High Overall Mortality. *Gastroenterology*. **134**(7):2191.
- TOYODA, H., FUKUDA, Y., HAYAKAWA, T., TAKAMATSU, J. & SAITO, H., 1998. Effect of GB virus C/hepatitis G virus coinfection on the course of HIV infection in hemophilia patients in Japan. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **17**(3):209-13.
- TUCKER, T. J., SMUTS, H. E., EEDES, C., KNOBEL, G. D., EICKHAUS, P., ROBSON, S. C. & KIRSCH, R. E., 2000. Evidence that the GBV-C/hepatitis C virus is primarily a lymphotropic virus. *Journal of Medical Virology*, **61**(1):52-58.
- TUCKER, J. T. & SMUTS. E. H., 2000. GBV-C/HGV Genotypes: Proposed Nomenclature for Genotypes 1–5. *Journal of Medical Virology*, **62**(1):82–83.
- VAN DE LAAR, T., PYBUS, O., BRUISTEN, S., BROWN, D., NELSON, M., BHAGANI, S., VOGEL, M., BAUMGARTEN, A., CHAIX, M. L., FISHER, M., GOTZ, H., MATTHEWS, G. V., NEIFER, S., WHITE, P., RAWLINSON, W., POL, S., ROCKSTROH, J., COUTINHO, R., DORE, G. J., DUSHEIKO, G. M. & DANTA, M., 2009. Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. *Gastroenterology*, **136**(5):1609-1617.
- VAN DER BIJ, A. K., KLOOSTERBOER, N., PRINS, M., BOESER-NUNNINK, B., GESKUS R. B., LANGE, J. M., COUTINHO, R. A. & SCHUITEMAKER, H., 2005. GB virus C co-infection and HIV-1 disease progression: The Amsterdam Cohort Study. *Journal of Infectious Diseases*, **191**(5):678–85.
- VIAZOV, S., RIFFELMANN, M., KHOUDYAKOV, Y., FIELDS, H., VARENHOLZ, C. & ROGGENDORF, M., 1997. Genetic heterogeneity of hepatitis G virus isolates from different parts of the world. Review. *Journal of General Virology*, **78**(3):577–581.

- VOIRIN, N., TRÉPO, C., ESTÈVE, J., CHEVALLIER, P., RITTER, J., FABRY, J. & VANHEMS, P., 2002. Effects of co-infection with hepatitis C virus and GB virus C on CD4 cell count and HIV-RNA level among HIV-infected patients treated with highly active anti-retroviral therapy. *AIDS*, **16**(11):1556–9.
- WANG, H. L., HOU, Y. D. & JIN, D. Y., 1997. Identification of a single genotype of hepatitis G virus by comparison of one complete genome from a healthy carrier with eight from patients with hepatitis. *Journal of General Virology*, **78**(12):3247–3253.
- WATANABE, M. A E., MIQUELÃO, A. K., CORAL DE OLIVEIRA, E. C., OLIVEIRA, K. B., NASSER, T. F., AOKI, M. N., SUZUKI, P.S., AMARANTE, M. K., VENÂNCIO, E. J. & ELBENS, M.M., 2008. Detection of GBV-C/HGV RNA in Cervico-Vaginal Smears from Healthy Individuals. *Brazilian Archives of Biology and Technology*, **51**(5):917-922.
- WEJSTAL, R., MANSON, A. S., WIDELL, A. & NORKRANS, G., 1999. Perinatal transmission of hepatitis G virus (GB virus type C) and hepatitis C virus infections--a comparison. *Clinical Infectious Diseases*, **28**(4):816-821.
- WILLIAMS, C. F., KLINZMAN, D., YAMASHITA, T. E., XIANG, J., POLGREEN, P. M., RINALDO, C., LIU, C., PHAIR, J., MARGOLICK, J. B., ZDUNEK, D., HESS, G. & STAPLETON, J. T., 2004. Persistent GB virus C infection and survival in HIV-infected men. *New England Journal of Medicine*, **350**(10):981-90.
- WILLIAMS, R. & RIORDAN, M. S., 2000. Acute liver failure: Established and putative hepatitis viruses and therapeutic implications. Review. *Journal of Gastroenterology and Hepatology*, **15**, Suppl:G17–G25
- WIWANITKIT, V., 2006. Hepatitis G, hepatitis SEN, hepatitis TT and hepatitis TT-like viruses: New emerging hepatitis viruses in pediatric patients. *Journal of Pediatric Infectious Diseases*, **1**(2):83-88.
- WOROBAY, M. & HOLMES, C. E., 2001. Homologous Recombination in GB Virus C/Hepatitis G Virus. *Molecular Biology and Evolution*. **18**(2):254–261.
- XIANG, J., GEORGE, S. L., WUNSCHMANN, S., CHANG, Q., KLINZMAN, D. & STAPLETON, J. T., 2004. Inhibition of HIV-1 replication by GB virus C infection through increases in RANTES, MIP- 1alpha, MIP-1beta, and SDF-1. *Lancet*, **363**(9426):2040-2046.
- XIANG, J., WÜNSCHMANN, S., DIEKEMA, D. J., KLINZMAN, D., PATRICK, K. D., GEORGE, S. L., STAPLETON, J. T., 2001. Effect of co-infection with GB virus C (hepatitis G virus) on survival among patients with HIV infection. *New England Journal of Medicine*, **345**(10):707-714.
- XIANG, J., KLINZMAN, D., MCLINDEN, H. J., SCHMIDT, N. W., LABRECQUE, R. D., GISH, R. & STAPLETON, T. J., 1998. Characterization of Hepatitis G Virus (GB-C Virus)

Particles: Evidence for a Nucleocapsid and Expression of Sequences Upstream of the E1 Protein. *Journal of Virology*, **72**(4):2738–2744.

XIANG, J., MCLINDEN, H. J., CHANG, Q., JORDAN, L. E. & STAPLETON, J. T., 2008. Characterization of a Peptide Domain within the GB Virus C NS5A Phosphoprotein that Inhibits HIV Replication. *Public Library of Science ONE*, **3**(7):e2580.

XIANG, J., MCLINDEN, J. H., CHANG, Q., KAUFMAN, T. M. & STAPLETON, J. T., 2006. An 85-aa segment of the GB virus type C NS5A phosphoprotein inhibits HIV-1 replication in CD4⁺ Jurkat T cells. *Proceedings of the National Academy of Science, USA*, **13**(42): 15570–15575

XIANG, J., SATHAR, A. M., MCLINDEN, H. J., LKINZMAN, D., CHANG, Q., & STAPLETON, J. T., 2005. South African GB Virus C Isolates: Interactions between Genotypes 1 and 5. *Journal of Infectious Diseases*, **192**(12):2147-2151.

XIANG, J., WUNSCHMANN, S., SCHMIDT, W., SHAO, J. & STAPLETON, J. T., 2000. Full-Length GB Virus C (Hepatitis G Virus) RNA Transcripts Are Infectious in Primary CD4-Positive T Cells. *Journal of Virology*, **74**(9):9125–9133.

YAN, J., CHEN, L.L., LUO, H.Y., MAO, Y.F. & HE, M., 2001. High frequencies of HGV and TTV infections in blood donors in Hangzhou. *World Journal of Gastroenterology*, **7**(5):637-641.

YAN, J. & DENNIN, H. R. 2000. A high frequency of GBV - C/HGV coinfection in hepatitis C patients in Germany. *World Journal of Gastroenterology*, **6**(6):833-841.

YANG, J. F., DAI, C. Y., CHUANG, W. L., LIN, W. Y., LIN, Z. Y., CHEN, S. C., HSIEH, M. Y., WANG, L. Y., TSAI, J. F., CHANG, W. Y. & YU, M. L., 2006. Prevalence and clinical significance of HGV/GBV-C Infection in patients with chronic hepatitis B or C. *Japanese Journal of Infectious Diseases.*, **59**(1):25-30.

YEO, A. E., MATSUMOTO, A., HISADA, M., SHIH, J. W., ALTER, H. J. & GOEDERT, J. J., 2000. Effect of hepatitis G virus infection on progression of HIV infection in patients with hemophilia. Multicenter Hemophilia Cohort Study. *Annals of Internal Medicine*, **132**(12):959–63.

ZHANG, W., CHALONER, K., TILLMANN, H. L., WILLIAMS, C. F. & STAPLETON, J. T., 2006. Effect of early and late GB virus C viraemia on survival of HIV-infected individuals: a meta-analysis. *HIV Medicine*, **7**(3):173–180.

ZUCKERMAN, A. J., 1996. Alphabet of hepatitis viruses. *Lancet*, **347**(9001):558–559.

ANEXOS



ANEXO I

TABELA ANEXO I. Dados epidemiológicos relativos às 87 amostras analisadas, onde foi detectada virémia de GBV-C. Encontram-se indicado, o local de recolha das amostras (AS- Hospital Fernando Dr. Fernando Fonseca, Amadora-Sintra; ou Taipas - Centro de Atendimento a Toxicodependentes das Taipas), a nacionalidade (A- Angolana; B- Brasileira; CV- Cabo Verdiana; F- Francesa; G- Guineense; M- Moçambicana; P- Portuguesa; SA- Sul Africana; ST- São Tomense), origem racial (cauc.- caucasiana; negra ou mista), sexo e idade dos indivíduos correspondentes. Estão também identificados os indivíduos que reportaram seropositividade para os vírus HIV (do tipo 1, 2 ou infecção mista) e HCV, aqueles que adquiriram o HIV por contacto sexual e os utilizadores de drogas injectáveis (UDI's). Relativamente aos indivíduos seropositivos para HIV, com base em estudos anteriores (Esteves *et al.*, 2002; Esteves *et al.*, 2003; Parreira *et al.*, 2005), encontra-se disponível para algumas amostras, a informação relativa à contagem de células CD4⁺, carga viral e subtipo HIV-1.

ID Amostras clínicas	Sexo	Idade	Origem de recolha	Nacionalidade e origem racial	Grupo de Risco			HIV	Subtipo HIV-1	Contagem de células CD4 ⁺	Carga Viral	HCV
					Infecção com HIV por contacto sexual							
					UDIs	Heterossexual	Homossexual					
0065	F	27	AS	P/cauc.	✓			✓		329	478	✓
0123	F	29	AS	A/negra		✓		✓	J	213	19095	
0594	M	25	AS	G/negra		✓		✓		338	50	
1017	M	29	AS	B/cauc.		✓		✓	B	438	140454	
1132	F	30	AS	P/cauc.	✓			✓	B	264	146	✓
1544	M	35	AS	P/cauc.	✓			✓	G	518	20344	✓
1619	F	31	AS	ST/negra		✓		✓	B	329	50110	
1631	M	24	AS	P/cauc.	✓			✓	G	420	6270	✓
1641	M	29	AS	P/cauc.	✓			✓	B	237	67057	✓
1642	F	27	AS	P/cauc.	✓			✓	B	520	29252	✓
2695	M	31	AS	P/cauc.				✓	G	70	23225	
3013	M	23	AS	P/cauc.	✓			✓	B	443	58008	✓
3273	F	37	AS	B/cauc.	✓			✓	B	136	79	✓
3423	M	30	AS	P/cauc.	✓			✓	B	622	29878	✓
3638	M	31	AS	P/cauc.	✓			✓	B	350	5258	✓
3806	F	20	AS	P/cauc.	✓			✓	G	1021	26464	✓
4065	F	24	AS	CV/negra		✓		✓		791	50	
4088	F	20	AS	P/cauc.		✓		✓	B	193	9977	
4135	M	33	AS	P/cauc.	✓			✓	B	444	78213	
5603	M	34	AS	A/negra		✓		✓	G	366	10198	
5658	M	34	AS	P/cauc.	✓			✓	G	361	3685	✓
6004	M	46	AS	P/cauc.	✓			✓	B	500	20042	✓
6073	F	27	AS	A/negra		✓		✓	D	520	2283	
6388	M	43	AS	CV/negra		✓		✓	B	115	59459	
6482	F	23	AS	P/cauc.	✓			✓	B	140	3714	✓
6821	F	22	AS	B/negra		✓		✓		349	830	
7251	M	30	AS	P/cauc.	✓			✓	G	727	16751	✓
7301	F	24	AS	P/cauc.		✓		✓	B	280	3403	
7642	F	23	AS	P/cauc.		✓		✓	B	412	49329	✓
7857	F	24	AS	M/cauc.		✓		✓	B	479	1315	
7888	M	45	AS	CV/negra		✓		✓	G	200	24993	
7893	M	45	AS	P/cauc.	✓			✓	B	199	39847	✓
8227	F	36	AS	P/cauc.	✓			✓	B	619	49632	✓
8395	F	32	AS	A/negra		✓		✓	G	262	4766	
8731	M	21	AS	P/cauc.	✓			✓	B	302	6892	✓

Continuação na página seguinte

ID Amostras clínicas	Sexo	Idade	Origem	Nacionalidade e origem racial	Grupo de Risco			HIV	Subtipo HIV-1	Contagem de células CD4 ⁺	Carga Viral	HCV
					Infecção com HIV por contacto sexual							
					UDIs	Heterossexual	Homossexual					
8919	M	29	AS	P/cauc.	✓			✓	B	278	50	✓
9295	F	17	AS	P/negra		✓		✓	G	971	407	
9317	M	33	AS	P/cauc.	✓			✓	D	14	106294	✓
9321	M	27	AS	P/cauc.	✓			✓	B	28	145102	✓
9363	M	34	AS	P/cauc.			✓	✓	B	1076	53422	
9456	M	32	AS	P/cauc.	✓			✓	G	567	4607	✓
9584	M	27	AS	P/negra		✓		✓	G	665	72240	
9709	M	31	AS	CV/negra	✓			✓	B	167	59589	✓
2102	M	30	Taipas	A/cauc	✓			✓	A			✓
2106	M	30	Taipas	M/negra	✓			✓				✓
4762	M	31	Taipas	A/misto	✓			✓	B			✓
5549	F	25	Taipas	A/mista	✓			✓	B			✓
5	M	42	Taipas	P	✓							
6	F	37	Taipas	P	✓							
7	M	32	Taipas	P	✓							
9	M	33	Taipas	P	✓							
10	M	34	Taipas	P	✓							✓
17	M	41	Taipas	P	✓							✓
20	M	50	Taipas	P	✓							✓
21	M	30	Taipas	A	✓							✓
28	F	32	Taipas	P	✓			✓				✓
29	F	34	Taipas	P	✓							
31	M	30	Taipas	P	✓							✓
38	M	36	Taipas	P	✓			✓				✓
40	M	49	Taipas	P	✓							✓
42	M	44	Taipas	P	✓			✓				✓
44	M	55	Taipas	P	✓			✓				✓
53	F	34	Taipas	P	✓							✓
58	F	35	Taipas	F	✓							✓
61	M	38	Taipas	P	✓							✓
68	M	34	Taipas	P	✓							
72	F	44	Taipas	P	✓			✓				✓
74	F	39	Taipas	P	✓							
75	M	33	Taipas	P	✓							
77	M	38	Taipas	P	✓							✓
78	F	29	Taipas	P	✓							✓
79	M	55	Taipas	P	✓							
80	F	32	Taipas	P	✓							
82	F	23	Taipas	P	✓							
85	M	32	Taipas	A	✓			✓				✓
86	M	28	Taipas	P	✓							✓
90	M	51	Taipas	A	✓							✓
91	M	34	Taipas	P	✓							
92	M	34	Taipas	P	✓							
94	M	41	Taipas	SA	✓							✓
97	M	29	Taipas	P	✓							
98	M	29	Taipas	P	✓			✓				✓

Continuação na página seguinte

ID Amostras clínicas	Sexo	Idade	Origem	Nacionalidade e origem racial	Grupo de Risco			HIV	Subtipo HIV-1	Contagem de células CD4 ⁺	Carga Viral	HCV
					Infecção com HIV por contacto sexual							
					UDIs	Heterossexual	Homossexual					
101	F	57	Taipas	P	✓							
103	F	17	Taipas	P	✓							
104	—	—	Taipas	—	✓							
107	F	31	Taipas	P	✓			✓				✓
111	M	37	Taipas	P	✓							

ANEXO II

TABELA ANEXO II. Representação das amostras analisadas por *nested* e *heminested*-PCR relativamente às regiões genómicas de GBV-C em estudo.

ID Amostras clínicas	Fragmentos analisados (correspondência com a respectiva região genómica de GBV-C)		
	5'UTR/E1	E1/E2	NS5A/NS5B
	<i>heminested</i> -PCR	<i>nested</i> -PCR	<i>nested</i> -PCR
0065	✓	✓	✓
0123	✓	✓	✓
0594	✓	✓	✓
1017	✓	✓	✓
1619	✓	✓	✓
1631		✓	✓
1641	✓	✓	✓
1642	✓	✓	✓
2695	✓	✓	✓
3013	✓	✓	✓
3273	✓	✓	✓
3423		✓	✓
3638	✓	✓	
3806	✓	✓	✓
4065	✓	✓	✓
4088	✓	✓	✓
5603	✓	✓	✓
6004	✓	✓	
6073		✓	✓
6388	✓	✓	✓
6482		✓	✓
6821	✓	✓	✓
7251	✓	✓	✓
7301	✓	✓	✓
7642	✓	✓	✓
7857	✓	✓	✓
7888	✓	✓	
7893		✓	✓
8227		✓	✓
8395	✓	✓	✓
8731		✓	
8919		✓	✓
9295	✓	✓	✓
9317		✓	✓
9321		✓	✓
9363	✓	✓	✓
9456	✓	✓	✓
9584	✓	✓	✓
9709	✓	✓	✓
75		✓	✓
79		✓	✓
80		✓	✓
94		✓	✓
97		✓	✓
111		✓	✓

ANEXO III

TABELA ANEXO III. Correspondência entre as amostras biológicas analisadas e a sua designação nos ensaios de PCR em Tempo Real.

Ensaio de PCR em Tempo Real	
ID Amostras	
A	9584
B	6073
C	111
D	75
E	79
F	9709
G	7642
H	7857
I	3068

ANEXO IV

COMPOSIÇÃO DOS MEIOS E SOLUÇÕES UTILIZADOS

Meio LB:

10g triptona

5g extracto de levedura

10g NaCl

15g agar*

Adicionar água para 1000ml e autoclavar

* Não adicionar para LB líquido

Solução de agarose (1%) (p/v) (50 ml):

0,5g agarose

50ml TAE 0,5X

Solução de agarose (1,5%) (p/v) (50 ml):

0,75g agarose

50ml TAE 0,5X

Solução de agarose (2%) (p/v) (50 ml):

1g agarose

50ml TAE 0,5X

Tampão de aplicação (em gel de agarose)

0,25%(p/v) azul de bromofenol

40%(p/v) sacarose

Tampão da enzima T4 DNA ligase 2X (Promega EUA):

60mM Tris-HCl, pH 7,8

20mM MgCl₂

20mM DTT

2mM ATP

10%(v/v) PEG

TSS:

LB 1X

10%(v/v) PEG 6000

5%(v/v) DMSO

50mM MgSO₄

(pH 6,5)

ANEXO V

TABELA ANEXO V. Sequências de referência utilizadas na análise filogenética. Encontra-se representado o número de acesso, (disponível nas bases de dados, EMBL/GenBank/DBJ), a origem geográfica dos isolados e o respectivo genótipo (G).

Seq. Referência de genoma completo de GBV-C	ORIGEM GEOGRÁFICA	G	Seq. de Referência de genoma parcial (região E2)	ORIGEM GEOGRÁFICA	G	Seq. de Referência de genoma parcial (região NS5B)	ORIGEM GEOGRÁFICA	G
AY196904.1	EUA	2	AB000164.1	EUA	3	U96126.1	EUA	2
U94695.1	China	3	AB000158.1	EUA	3	U96125.1	EUA	2
AF121950.1	EUA	2	AB000150.1	EUA	3	U96124.1	EUA	2
AB003291.1	Japão	1	AY951980.1	África do Sul	1	U96123.1	EUA	2
NC_001710.1	EUA	2	AY951979.1	África do Sul	5	U96122.1	EUA	2
AF081782.1	China	2	AY951978.1	África do Sul	5	U96121.1	EUA	2
AB003290.1	Japão	3	AY951977.1	África do Sul	5	U96120.1	EUA	3
AB003288.1	Japão	3	AY951976.1	África do Sul	5	U96119.1	EUA	3
D87255.1	Japão	2	AY951975.1	África do Sul	1	U96118.1	EUA	3
D90601.1	Japão	3	AY951974.1	África do Sul	2	U96117.1	EUA	3
D90600.1	Japão	2	AY951973.1	África do Sul	1			
U45966.1	EUA	2	AY951972.1	África do Sul	1			
AB003289.1	Japão	2	AY951971.1	África do Sul	5			
AB003292.1	Indonésia	6	AY951970.1	África do Sul	1			
AB003293.1	Japão	3	AY951969.1	África do Sul	1			
AB008335.1	Japão	3	AY951968.1	África do Sul	5			
AB008342.1	Japão	3	AY951967.1	África do Sul	5			
AB013500.1	Ghana	1	AY951966.1	África do Sul	5			
AB018667.1	Vietnam	4	AY951965.1	África do Sul	5			
AF006500.1	China	3	AY951964.1	África do Sul	5			
AF031829.1	EUA	2	AY951963.1	África do Sul	5			
AF104403.1	França	2	AY951962.1	África do Sul	5			
AF309966.1	Alemanha	2	AY951961.1	África do Sul	1			
D87263.1	Japão	3	AY951960.1	África do Sul	5			
D87708.1	Japão	3	AY951959.1	África do Sul	5			
D87709.1	Japão	3	AY951958.1	África do Sul	1			
D87710.1	Japão	3	AY951957.1	África do Sul	5			
D87711.1	Japão	3	AY951956.1	África do Sul	5			
D87712.1	Japão	3	AY951955.1	África do Sul	5			
D87713.1	Japão	3	AY951954.1	África do Sul	2			
D87714.1	Japão	3	AY951953.1	África do Sul	5			
D87715.1	Japão	3	U87664.1	EUA	2			
U36380.1	Gana	1	U87663.1	Jamaica	2			
AY949771.1	África do Sul	5	U87662.1	Jamaica	2			
U63715.1	Egipto	2	U87661.1	Egipto	2			
			U87660.1	Egipto	2			
Recombinantes			U87659.1	EUA	1			
U75356.1	China		U87656.1	Egipto	2			
AB013501.1	Bolívia		U87655.1	Inglaterra	2			
AB021287.1	Mianmar		U87654.1	Egipto	2			
			U87653.1	Grécia	2			
			AF017533.1	China	3			
			AF017532.1	China	3			

ANEXO VI

ABSTRACT

Prevalence of infection and genetic analysis of GB virus C in a population group comprising a large number of intravenous drug users (IDUs). CRISTINA BRANCO¹, AIDA ESTEVES^{1, 2}, JOÃO PIEDADE^{1, 2}, AND RICARDO PARREIRA^{1, 2}.

¹Unidade de Ensino e Investigação Virologia, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (UNL); ²Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa (UNL).

Key words: GBV-C/HGV, intravenous drug users, E2 genotyping, phylogenetic analysis.

GB virus C (GBV-C) (also known as hepatitis G virus or HGV) is a tentative member of the *Flaviviridae* family, closely related to the hepatitis C virus (HCV). Although initially regarded as a potential cause of hepatitis, more recent studies have found no evidence for association between GBV-C and human disease. GBV-C is very common among individuals infected with the human immunodeficiency virus (HIV) and/or HCV, and among IDUs, most probably due to similar transmission routes. Due to the scarcity of information regarding GBV-C infection in Portugal, we decided to estimate its prevalence in a group of 214 individuals from the Lisbon metropolitan area, including a large number of IDUs (70.6%), and to carry out a genetic characterization of the circulating viral strains. GBV-C viremia was assessed by nested PCR amplification of a section of the conserved 5'-untranslated region using total RNA extracted from plasma samples. The genetic characterization of GBV-C strains was carried out on a randomly selected group of 44 viral strains, focusing on the analysis of a genomic PCR fragment comprising most of the E2 surface glycoprotein coding region.

The overall prevalence of GBV-C infection was 40.6%. Among those with GBV-C viremia, monoinfection was detected only in 23.0% of the samples, while coinfection with HIV, HCV, or HIV/HCV was revealed in 19.5%, 16.1%, and 41.4% of the population, respectively. Statistically significant associations were detected between GBV-C viremia and the age (≥ 25 years) or the use of intravenous drugs. Phylogenetic analysis of the E2 region, established segregation of Portuguese GBV-C strains with genotypes 1 (G1, n=10) and 2 (G2, n=34) reference sequences, irrespective of the analytic method used (Neighbor Joining, Maximum Likelihood or a Bayesian Markov chain Monte Carlo approach). Unexpectedly, 15 strains assigned to G2 were shown to form a separate cluster in all phylogenetic trees. This group of sequences, designated G2*, was characterized by the overall genetic conservation (7.8% diversity) and the presence of specific amino acid polymorphisms (Q134, H258, and F182), when compared to its sister clade (G2). However, different from what the phylogenetic tree topologies might, at first, suggest, analysis of the distributions of intra and inter genotype pairwise genetic distances did not support the appointment of G2* as a new GBV-C genotype, but rather as a geographic subcluster within the G2 radiation.

INTRODUCTION

GB virus C (GBV-C) is a tentative member of the *Flaviviridae* family, closely related to the hepatitis C virus (HCV).¹ This virus exists as a collection of genotypes with a global distribution that correlates with the geographical origin of their hosts.² Although initially regarded as a potential cause of hepatitis, justifying its former designation as hepatitis G virus, or HGV, more recent studies have found no evidence for association between GBV-C and human disease.³ On the contrary, GBV-C has been associated with prolonged AIDS-free survival, higher CD4⁺ cell counts, and lower HIV viral loads in persons coinfected with HIV and GBV-C.⁴ GBV-C infection is relatively frequent among humans, with up to 2% of the blood donors in the USA showing active viral replication. However, the rates of infection with GBV-C among individuals at high risk of exposure to parenterally transmitted viruses, e.g. intravenous drug users, may be significantly higher.⁵ Due to the scarcity of information regarding GBV-C infection in Portugal, we decided to estimate its prevalence in a group of individuals from the Lisbon metropolitan area, including a large number of IDUs (70.6%), and to carry out a genetic characterization of the circulating viral strains.

MATERIALS AND METHODS

SUBJECTS

This study was based on a convenience sample which included 214 individuals living in the Lisbon metropolitan area (63.1% males, 36.9% females), aged 17 to 67 years (mean 35.2 years). A considerable number of these individuals (70.6%) reported use of intravenous drugs, and the prevalence of infection with HIV or HCV was found to be high (71.0% and 52.3%, respectively).

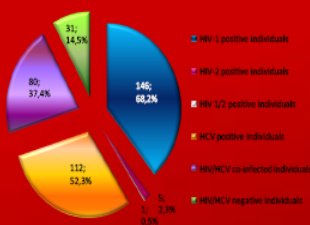


Figure 1. Studied population. The number of samples corresponding to each studied group, as well as respective percentage, are indicated.

1- GBV-C VIREMIA

Viral RNA was extracted from 150µl of plasma using the Instant Virus RNA kit (Analytik Jena, Germany) and the presence of active viremia was screened by RT-PCR (nested) using the highly conserved 5' untranslated region (UTR) as a target. A specific amplicon with approximately 250bp, was amplified using PCR primers UTRForO/E1rev2 (5'-TGTAAGCCACTATAGGTGGTCTTAAG-3'/5'-AAATGCCACCGCCCTACCGAAGG-3', 1st round) and UTRForO/E1rev1 (5'-GGCAGCGTCCACAGGTGTGGCCCTAC-3'/5'-GGGCTGGTGGCCATGCTTTTCC-3', 2nd round) (Figure 2).

2- E1/E2 ENVELOPE GLYCOPROTEIN AMPLIFICATION

The genetic characterization of circulating GBV-C strains was carried out on a randomly selected group of 44 samples where viremia had been previously demonstrated. For that matter we analysed the nucleotide and predicted protein sequences of a viral genomic segment corresponding to part of the E1/E2 envelope glycoprotein coding region. This was obtained as a 1.1 kb fragment via RT-PCR (nested), using primer pairs E2For1/E2rev3 (5'-AGTGAGTTTGGAGATGGAAGCAAGC-3'/5'-GGGAAWGCYCCCGAGCGAGCTCCAC-3', 1st round) and E2For2/E2rev2 (5'-GTGTGATGCGCCCAAGGCGCTGGGT-3'/5'-CCACARACRAGAACATBAGGCGTG-3', 2nd round).



Figure 2. Schematic representation of the GBV-C genome structure. The primers used for the amplification of 5' UTR and E1/E2 glycoprotein regions by nested PCR, are indicated. The approximate size of the obtained amplicons is also illustrated.

3- PHYLOGENETIC ANALYSIS

Sequences of partial E1/E2 genomic region (864nt) were obtained and aligned (MAFFT vs. 6.) with the corresponding region from full-length GBV-C genome sequences available in the public domain (a total of 62 references, representing the 6 GBV-C genotypes). Three alternative analytic methods were used for the construction of phylogenetic trees (Neighbor Joining, Maximum Likelihood or a Bayesian Markov chain Monte Carlo approach). Neighbor-joining (NJ) trees, obtained using MEGA vs. 4, were based on genetic distance matrices corrected with the Kimura 2-parameter formula (K2P). Alternatively, the Bayesian Markov chain Monte Carlo approach was run for 6x10⁶ generations under a GTR model. Finally, the Maximum Likelihood trees, were obtained using PALP 4.0b10, and nucleotide rate heterogeneity was estimated using a γ distribution for the variable sites using Modeltest.

4- AMINO ACID ANALYSIS

The ratio of non-synonymous to synonymous amino acid substitutions (dN/dS) was estimated according to the method described by Nei and Gajbordi using the SNAP tool. The degree of variability of each amino acid position in multiple alignments of the putative GBV-C amino acid sequences, was calculated using Entropy (Shannon entropy-two option). Both bioinformatic tools are available at <http://www.hiv.lanl.gov/content/sequence/HIV/HIVTools.html>.

5- STATISTICAL ANALYSIS

Statistical analyses included, when appropriate, P values of the χ^2 , χ^2 with Yates correction on the two-tailed Fisher's exact tests. Values <0.05 were considered statistically significant.

RESULTS

Prevalence of GBV-C active replication demonstrated by detection of viremia

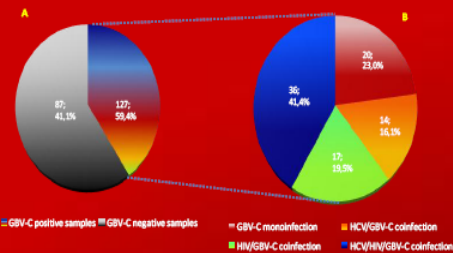


Figure 3.

GBV-C infection rate. **A)** A specific emphasis with the expected size was detected in 87 of the 214 samples analyzed. **B)** Among those with GBV-C viremia, GBV-C infection rate is also indicated for GBV-C monoinfection and coinfection with HIV, HCV, and HIV/HCV, in absolute values and corresponding percentage.

Phylogenetic analysis

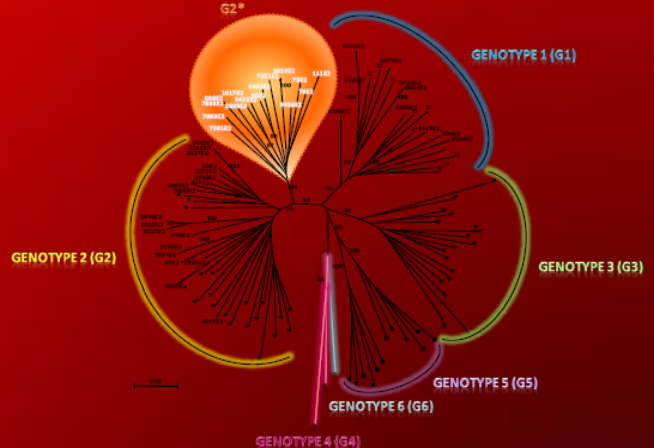


Figure 4. Phylogenetic analysis of GBV-C sequences using K2P genetic distances/NJ (>75% bootstrap values were considered). References used include genotype 1 (), 2 (), 3 (), 4 (), 5 () and 6 () sequences.

Analysis of a subtype 2 GBV-C sub-cluster (G2*)

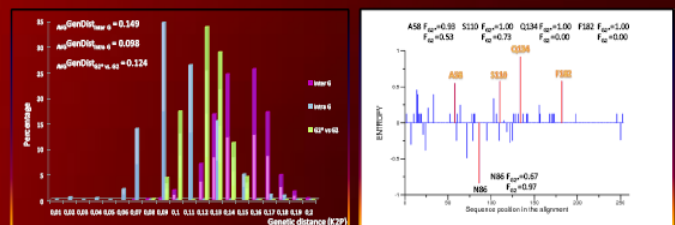


Figure 5.

Distribution of intra-genotype (intra G) and inter-genotype (inter G) pairwise genetic distances, corrected with the K2P formula. The distribution of the pairwise genetic distance values when G2* were compared to G2 is represented by the green bars. $F_{0.05}$ indicates average genetic distance values.

Figure 6.

Graphical representation of entropy differences between the putative G2 and G2* amino acid sequences. Residues which were found to be highly conserved in G2 and G2* (P values <0.005) are indicated along with their respective frequencies in each data set ($F_{0.05}$, $F_{0.10}$).

CONCLUSIONS

* A high rate of GBV-C infection was detected (40.6%). Double HCV/GBV-C and HIV/GBV-C, as well as triple HCV/HIV/GBV-C infections were frequent. Statistically significant associations were detected between GBV-C viremia and the age (≥ 25 years of age, $P=0.03$) or the use of intravenous drugs ($P=0.04$) (Figure 3).

* All phylogenetic inference methods yielded trees with similar topologies for the major clades, and established segregation of Portuguese GBV-C strains with genotypes 1 (G1, n=10) and 2 (G2, n=34) reference sequences (Figure 4).

* A group of 15 strains assigned to G2, designated by G2*, were shown to form a separate cluster in all phylogenetic trees (orange bubble in Figure 4).

* Analysis of the distributions of intra and inter genotype pairwise genetic distances support the appointment of G2* as a geographic subcluster within the G2 radiation, rather than a new GBV-C genotype (Figure 5).

* The G2* cluster was characterized by the overall genetic conservation (7.8% amino acid diversity) and the presence of specific amino acid polymorphisms (A58, S110, Q134 and F182), when compared to its sister clade (G2), as demonstrated by entropy (Shannon index) analysis (Figure 6).

REFERENCES

1. Barry TR et al. 1996. Sequence and genomic organization of GBV-C: a novel member of the flaviviridae associated with human non-A-E hepatitis. *J Med Virol* 48:80-87.
2. Muerhoff S et al. 2005. African origin of GB virus C determined by phylogenetic analysis of a complete genotype 5 genome from South Africa. *Journal of General Virology* 86, 1729-1735.
3. Sather AN et al. 2000. GB virus C/hepatitis G virus (GBV-C/HGV): still looking for a disease. *Review. Int J Exp Path* 81, 305-322.
4. George LS and Naiman D. 2005. What You Need to Know About GB Virus C. *Current Gastroenterology Reports* 7:54-62.
5. Baggio-Zappia LG and Granato HFC. 2009. Review. HIV/GB virus C co-infection: an overview. *Int J Clin Chem Lab Med* 47(1):112-119.

ACKNOWLEDGMENTS

This study is part of a larger research project involving the genetic characterization of blood-borne viruses circulating in the injecting drug users of the Lisbon metropolitan area, and was partially funded by Fundação GlaxoSmithKline dos Clínicos de Saúde, Portugal. We would like to thank the staff of Centro de Atendimento de Toxicodependentes das Taipais, Av do Brasil 53, Lisbon, for their kind cooperation, and Teresa Venancio for technical assistance.

ANEXO VII

A New Genotype 2 Subcluster Identified Among GBV-C Strains Circulating in the Lisbon Metropolitan Area of Portugal

Cristina Branco, Aida Esteves, João Piedade, and Ricardo Parreira*

Virology Unit / Parasitology and Medical Microbiology Unit, Institute of Hygiene and Tropical Medicine, New Lisbon University (UNL), Lisboa, Portugal

The rate of infection by the GBV-C virus was investigated in a group of 214 individuals at high risk of infection with parenterally transmitted viruses, and all living in the Lisbon metropolitan area (Portugal). RNA was extracted from plasma samples, and a fragment of the 5'-UTR was amplified by RT-PCR, disclosing a high prevalence of infection (40.7%). Most probably due to similar modes of viral transmission, the majority of GBV-C (+) individuals were found to be coinfecting with HIV and/or HCV. A genomic region covering part of the E1/E2 glycoprotein coding sequence was amplified from approximately half of the GBV-C positive samples (44/87). Phylogenetic analysis of nucleotide sequences showed segregation of Portuguese GBV-C strains with genotype 1 (G1, $n = 10$) and genotype 2 (G2, $n = 24$) references. Genotype 1 was significantly associated with the African descent of those infected. Curiously, some of the strains assigned to genotype 2 were shown to form a separate cluster (designated G2*) in both neighbor-joining and Bayesian phylogenetic trees, which was confirmed by multivariate principal coordinate analysis. However, analysis of the distribution of intra- and intergenotype genetic distances support the hypothesis that rather than corresponding to a new viral genotype, G2* is a geographical subcluster within the genotype 2 radiation. **J. Med. Virol. 82:452–459, 2010.** © 2010 Wiley-Liss, Inc.

KEY WORDS: GBV-C; genotyping; genetic variability; phylogenetic analysis

INTRODUCTION

The GBV-C virus is a member of the *Flaviviridae* family, closely related to GBV-A and GBV-B, and sharing 25–30% amino acid sequence similarity with the hepatitis C virus (HCV) [Leary et al., 1996].

However, despite a similar genomic organization and phylogenetic proximity to hepaciviruses, GBV-C appears to lack a complete nucleocapsid protein gene, the 3'-untranslated region (UTR) displays a less complex organization, the E2 protein lacks the hyper-variable region at the N-terminus, and unlike HCV may establish a persistent infection in the absence of seroconversion to anti-E2 [George and Varmaz, 2005].

GBV-C is classified into five major groups (or genotypes) differing in 11–14% in their polyprotein coding sequence [Smith et al., 1997, 2000]. Nevertheless, the existence of a sixth genotype has already been proposed by Muerhoff et al. [2006], as revealed by the analysis of 5'-UTR sequences. These viral genotypes display an overall geographical distribution which mirrors that of the human racial groups. This has led to the suggestion that GBV-C may have infected the ancestral human population in Africa, its consequent diversification reflecting human migrations [Pavesi, 2001; Muerhoff et al., 2005].

Numerous clinical, epidemiological, and cellular biology studies argue against the initial association of GBV-C with viral hepatitis [reviewed in George et al., 2002], which led to its alternative designation, in the past, as hepatitis G virus (or HGV). Curiously, not only does GBV-C not seem to be a human pathogen, but it is also associated with prolonged AIDS-free survival, higher CD4⁺ cell counts, and lower HIV viral loads in persons coinfecting with HIV and GBV-C [reviewed in Shankar et al., 2008].

Additional supporting information may be found in the online version of this article.

Grant sponsor: Fundação GlaxoSmithKline das Ciências da Saúde, Portugal.

*Correspondence to: Ricardo Parreira, Instituto de Higiene e Medicina Tropical (IHMT/UNL), Unidade de Virologia, Rua da Junqueira 100, P-1349-008 Lisboa, Portugal.
E-mail: ricardo@ihmt.unl.pt

Accepted 13 October 2009

DOI 10.1002/jmv.21703

Published online in Wiley InterScience
(www.interscience.wiley.com)

GBV-C replicates at low levels in hepatocyte-derived cell lines, in accordance with its low hepatic titers in vivo [George and Varmaz, 2005; George et al., 2005]. Furthermore, since its genome can be readily detected in peripheral blood mononuclear cells [Kobayashi et al., 1999], bone marrow [Radkowski et al., 2000], and primary B and T lymphocytes [George et al., 2005], GBV-C is considered a lymphotropic virus [Baggio-Zappia and Hernandez Granato, 2009].

Infection with GBV-C is very common and may last for decades. Nearly 2% of the healthy blood donors are viremic for GBV-C, and in up to 17% the presence of E2 IgG indicates a previous infection [George and Varmaz, 2005; Reshetnyak et al., 2008]. However, the rates of infection with GBV-C among individuals at high risk of exposure to parenterally transmitted viruses such as HCV or HIV, for example, intravenous drug users, may be significantly higher [reviewed in George et al., 2002; Baggio-Zappia and Hernandez Granato, 2009].

In the current study, the prevalence of GBV-C viremia was determined in a group of individuals, all residents in the Lisbon metropolitan area and sharing a high risk of infection with parenteric viruses. A genetic analysis of the viral genomic region covering most of the envelope E2 glycoprotein coding sequence, amplified from a random sub-sample of the population studied, is also described.

MATERIALS AND METHODS

Subjects

This study was based on a convenience sample which included 214 individuals (63.1% males, 36.9% females), aged 17–67 years (mean 35.2 years), and residents in

the Lisbon metropolitan area (Portugal). This group of individuals shared a high risk of infection by parenterally transmitted viruses, either due to unprotected sex or injection of illegal drugs (reported by 70.6% of the individuals studied, Table I). Based on information from previous reports [Esteves et al., 2002, 2003] and ongoing studies, 71.0% were shown to be infected with HIV and 52.3% were infected with HCV (A. Esteves, unpublished work). Most individuals (67.8%) had Portuguese nationality, 28.5% were Africans (from Angola, Cape Verde, Guinea-Bissau, Mozambique, São Tomé and Príncipe, and South Africa), and 2.8% reported French, Brazilian, or Spanish citizenship. The nationality of two individuals was unknown.

RNA Extraction and Amplification of GBV-C Sequences

Viral RNA was extracted from 150 µl of plasma using the Instant Virus RNA kit (Analytik Jena, Jena, Germany) according to the manufacturer's instructions. The extracted RNA was kept at -80°C in 5–20 µl aliquots until further use. The RevertAidTM H Minus First Strand cDNA Synthesis kit and random hexaprimers (Fermentas, Vilnius, Lithuania) were used for the reverse transcription (RT) of 5 µl of the extracted viral RNA. The obtained cDNA served as template for the amplification of GBV-C sequences by nested polymerase chain reaction (nested PCR) using the PuRe Taq Ready-to-Go PCR Beads (GE Healthcare, Munich, Germany), as recommended by the supplier. The 5'-UTR region (approximately 250 bp) was amplified using PCR primers UTRforO/E1rev2 (5'-TGGTAGCCACTA-TAGGTGGGTCTTAAG-3'/5'-AAATGCCACCCGCCCT-

TABLE I. Demographic, Clinical, and Epidemiological Characteristics of the Individuals Studied

Number of individuals studied (n = 214)	GBV-C (+), n = 87	GBV-C (–), n = 127
Gender		
Male (63.1% of the total)	54 (62.1%)	81 (63.8%)
Female (36.9% of the total)	33 (37.9%)	46 (36.2%)
Age ^a	32.8 [17–57]	36.8 [21–67]
Nationality		
Portuguese (67.8% of the total)	63 (72.4%)	82 (64.5%)
African (28.5% of the total)	19 (21.9%)	42 (33.1%)
Other ^b (2.8% of the total)	4 (4.6%)	2 (1.6%)
Unknown (0.9% of the total)	1 (1.1%)	1 (0.8%)
History of injecting drug use (n = 151) (70.6% of the total)	68 (78.2%)	83 (65.4%)
Infection with HCV (n = 112) (52.3% of the total)	50 (57.5%)	62 (48.8%)
Infection with HIV (n = 152) (71.0% of the total)	55 (63.2%)	97 (76.4%)
Transmission		
Sexual	18 (32.7%)	42 (43.3%)
Use of injecting drugs	36 (65.5%)	53 (54.6%)
Unknown	1 (1.8%)	2 (2.1%)
CD4 ⁺ cell count ^{a,c} (n = 123)	4.0×10^2 [$14-1.1 \times 10^3$]	4.6×10^2 [$13-4.3 \times 10^3$]
Viral load ^{a,d} (n = 124)	3.0×10^4 [$<50-1.4 \times 10^5$]	3.9×10^4 [$<50-5.0 \times 10^5$]

The percentages calculated (gender, nationality, history of injecting drug use, infection with HCV and infection with HIV) are relative to the total number of GBV-C (+) or GBV-C (–) individuals (n = 87 and n = 127, respectively). The values referring to the different routes of HIV transmission (infection with HIV section) were calculated taking into account the total number of HIV/GBV-C (+) (n = 55) or HIV/GBV-C (–) (n = 97) individuals.

^aMean [range].

^bIndicates Brazilian, Spanish, and French nationality.

^cNumber of cells/µl.

^dViral RNA copies/ml.

CACCCGAAGG-3'; coordinates 160–186/509–534 in NC_001710) and UTRfor1/E1rev1 (5'-GCGCACGGTCC-ACAGGTGTTGGCCCTAC-3'/5'-GGRGCTGGGTGGCC-YCATGCWTTTCCC-3'; coordinates 224–251/450–476 in NC_001710) in the first and second rounds of PCR, respectively. In both amplifications, reaction conditions included an initial denaturation step at 95°C for 2 min, followed by 35 cycles of 95°C for 45 sec, 55°C for 45 sec, and 72°C for 45 sec. The E1/E2 envelope glycoprotein region was obtained as a 1.2 kbp fragment using primer pairs E2for1/E2rev3 (5'-AGTGAGTTTGGAGATGGA-CTGARCAG-3'/5'-GGGAAWGCYCCCCGAGCRAGC-TTCCAC-3'; coordinates 966–992/2456–2482 in NC_001710) and E2for2/E2rev2 (5'-GTGTGAYTGCCCC-AAYGGYCCCTGGGT-3'/5'-CCACARCACRAGRAAC-ATBAGGCGYTG-3'; coordinates 1295–1321/2433–2459 in NC_001710). PCR conditions included an initial denaturation step at 95°C for 2 min, followed by 25 cycles of 95°C for 45 sec, 55°C for 45 sec, and 72°C for 2.5 min, followed by 10 extra cycles with a rise in the hybridization temperature to 58°C, and a 5 sec increment in each extension step. Similar conditions were used in the first and second amplification rounds.

DNA Sequencing and Nucleotide/Amino Acid Sequence Analyses

PCR amplicons were purified directly from the reaction mixtures using the QIAquick PCR Purification kit (Qiagen, Valencia, CA), and directly sequenced using their respective second round PCR primers. Whenever necessary (low resolution of the sequencing reactions), the amplicons were cloned in *Escherichia coli* XL1-Blue MRF⁺ (Stratagene, La Jolla, CA) using pGEM-T Easy (Promega, Madison, WI) prior to DNA sequencing. Assembly of forward and reverse DNA sequences into single contigs was carried out using BioEdit 7.0.2 [Hall, 1999].

Phylogenetic inference analyses were based on the construction of nucleotide multiple sequence alignments with MAFFT vs. 6 [Kato and Toh, 2008]. Neighbor-joining (NJ) trees, obtained using MEGA vs. 4 [Tamura et al., 2007], were based on genetic distance matrixes corrected with the Kimura 2-parameter formula (K2P). The robustness of these phylogenetic trees was assessed by bootstrapping using 1,000 pseudo-replicates of the original sequence data. Alternatively, phylogenetic inference was based on a Bayesian Markov chain Monte–Carlo approach, run for 6×10^6 generations under a GTR model, and nucleotide rate heterogeneity was estimated using a γ distribution for the variable sites, using MrBayes v3.0b4 [Huelsenbeck et al., 2001].

The putative mosaic structure of GBV-C sequences was investigated using the SimPlot software [Lole et al., 1999], using genotype references obtained from the EMBL/GenBank/DBJ databases. Bootscanning analysis was carried out using NJ phylogenetic trees and K2P genetic distances estimated from an 180 bp

window moving along the sequences with 10 bp increments.

The ratio of non-synonymous to synonymous substitutions (dN/dS) was estimated according to the method described by Nei and Gojobori [1986] using the SNAP tool (<http://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html>). The degree of variability of each amino acid position in multiple alignments of the putative GBV-C amino acid sequences, evaluated on the basis of the Shannon entropy function, was calculated using Entropy (Shannon entropy-two option, <http://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy.html>). Finally, principal coordinate analysis was carried out using PCOORD (<http://www.hiv.lanl.gov/content/sequence/PCOORD/PCOORD.html>).

The coordinates of the amino acid residues mentioned in the text refer to those of the NC_001710 reference sequence. The nucleotide sequences reported in this study have been deposited at the EMBL/GenBank/DBJ sequence databases under accession numbers FN396832 to FN396875.

Statistical Analyses

Statistical analyses included, when appropriate, P values of the χ^2 , χ^2 with Yates correction or the two-tailed Fisher's exact tests. Values <0.05 were considered statistically significant. For the Shannon entropy-two analysis, statistical significance for the observed entropy differences was assessed following a Monte–Carlo randomization strategy (with replacement) where P equals the number of random samples ($n \leq 5$) that can have a higher entropy difference than the actual data set, in 1,000 randomizations ($P \leq 0.005$).

RESULTS

In this report the prevalence of GBV-C viremia was estimated in a group of 214 individuals at high risk of exposure to blood-borne viruses (mostly by unprotected sexual contact or use of illegal injected drugs; Table I). A genetic characterization of both nucleotide and putative amino acid sequences of a genomic region comprising most of the E2 glycoprotein gene was also carried out for a random selection of viral strains.

Detection of viremia was achieved by RT-PCR amplification of a highly conserved region of the viral genome with approximately 250 bp (coordinates 224–476 in NC_001710). For this purpose, primers complementary to 5'-UTR were designed using a multiple sequence alignment of 45 GBV-C complete genomes retrieved from public databases. A specific amplicon with the expected size was detected in 87 of the 214 samples analyzed, which corresponded to an estimated rate of viremia of 40.7% (Table II). Among those with GBV-C viremia, monoinfection was detected in only 23.0% of the samples ($n = 20$), while coinfection with HIV, HCV, and HIV/HCV was disclosed in 19.5% ($n = 17$), 16.1% ($n = 14$), and 41.4% ($n = 36$) of the samples, respectively (Table II). Infection with GBV-C was not found to be associated with the gender, origin

TABLE II. Rates of Infection With GBV-C, and/or HIV and/or HCV, and Distribution of GBV-C Genotypes in the Population Studied

GBV-C viremia (n = 87)	
GBV-C ^a	20 (23.0%)
HCV/GBV-C coinfection	14 (16.1%)
HIV/GBV-C coinfection	17 (19.5%)
HCV/HIV/GBV-C coinfection	36 (41.4%)
GBV-C genotypes (n = 44)	
Genotype 1	10 (22.7%)
Genotype 2	34 (77.3%)

^aMonoinfection with GBV-C.

(African vs. non-African), or CD4⁺ cell count (≤ 250 vs. >250 cells/ μ l) or HIV-1 viral load ($\leq 10^3$ vs. $>10^3$ RNA copies/ml) in persons coinfectd with both viruses. Nonetheless, statistically significant associations were detected between GBV-C viremia and the age (≥ 25 years old, $P = 0.03$) or the use of intravenous drugs ($P = 0.04$).

As shown previously, phylogenetic analyses of sequences covering GBV-C E2 coding region produce congruent trees with those obtained for the whole viral genome [Smith et al., 2000]. With this in mind, a genetic

characterization of the GBV-C strains was carried out in 44 randomly selected viral samples. Once again, a multiple sequence alignment of complete viral genomes was used to assist the design of two primer sets for the amplification of a DNA fragment of approximately 1.2 kbp (coordinates 1295–2459 in NC_001710). This amplicon included the 3'-terminal 20% of the E1 (approximately 220 bp) and 96% of the E2 coding sequences (approximately 940 bp).

The phylogenetic analysis of the Portuguese GBV-C sequences was based on a multiple sequence alignment consisting of a total of 62 references, representing the 6 GBV-C genotypes [as previously defined by Muerhoff et al., 2006]. The phylogenetic trees obtained from either K2P genetic distances using a clustering algorithm (NJ), or Bayesian approaches, are shown in Figure 1. Both methods yielded trees with essentially similar topologies, supported by high ($>75\%$) bootstrap values (Fig. 1A) or significant (0.90) posterior probabilities (Fig. 1B) for most major clades (especially those identifying the different viral genotypes). These results showed the segregation of the GBV-C strains circulating in the Lisbon metropolitan area with reference sequences from genotypes 1 (G1; n = 10), and 2 (G2; n = 34) (see Table II).

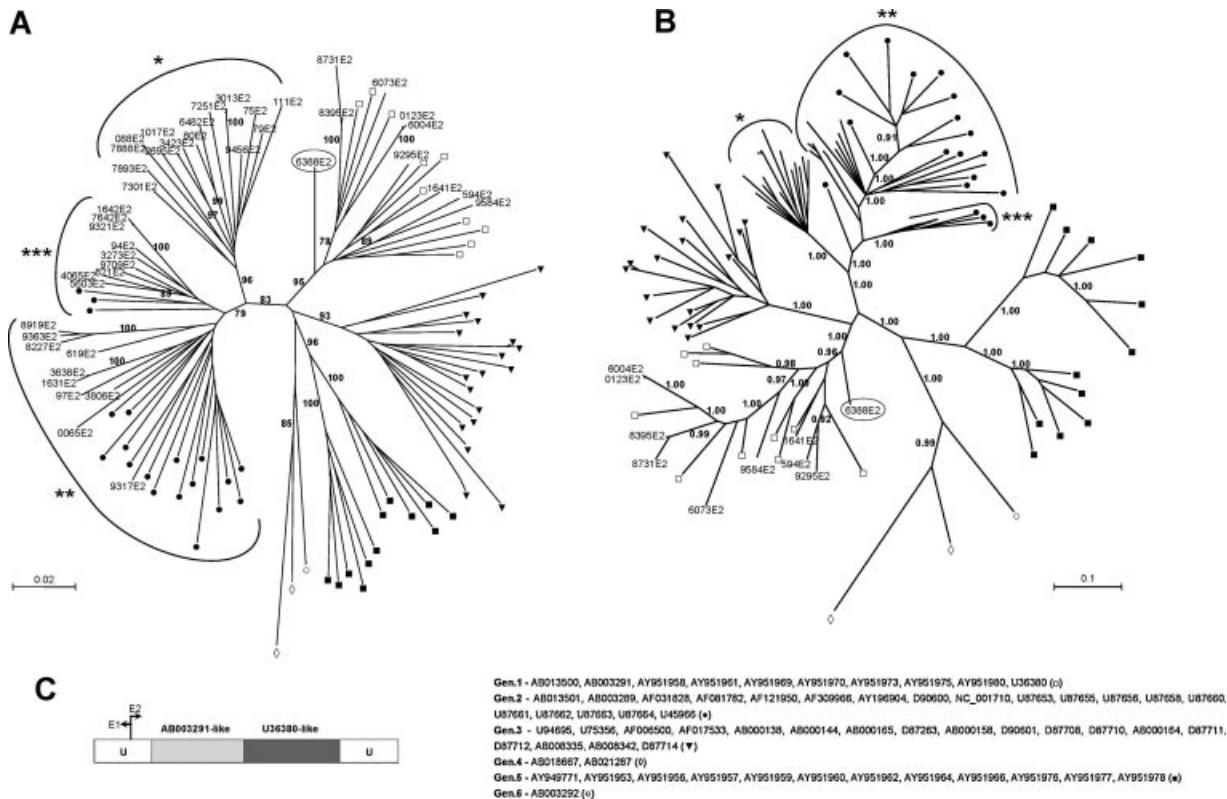


Fig. 1. Phylogenetic analysis of GBV-C sequences using K2P genetic distances/NJ (A) or Bayesian (B) approaches. The reference sequences used are indicated on the lower-right corner of the figure. Due to graphical constraints (especially in the Bayesian tree), groups of sequences have been designated as *, **, and ***. Again, due to graphical constraints, only the Portuguese Gen.1 GBV-C sequences are indicated in the Bayesian (B) phylogenetic tree. When appropriate, bootstrapping values ($>75\%$, in A) or posterior probabilities (>0.9 , in B)

are indicated. The putative genetic structure of the recombinant sequence 6388E2 (encircled in A and B), as defined by bootscanning analysis, is indicated in (C). The divergent arrow represents the approximate limits of the E1 and E2 coding sequences, and U indicates unknown genetic origin. The most closely related reference sequence to each of the intragenotype section of 6388E2 is indicated above the grey rectangles. The scale bars in (A) and (B) indicate 0.02 and 0.1 nucleotide substitutions per site, respectively.

Genotype 1 was found to be significantly associated with the African origin ($P=0.006$) of the individuals included in this study. One of the sequences analyzed (6388E2; encircled in Fig. 1A) segregated prematurely from all the others in the genotype 1 radiation, hinting to its possible recombinant or divergent nature. The first hypothesis was investigated by bootscanning analysis, which revealed that 6388E included a central intra-G1 recombinant region (Fig. 1C). However, the genetic origins of its 5' and 3'-ends could not be clearly defined, and were consequently designated as unknown (U).

Interestingly, 15 of the sequences analyzed segregated away from the remainder genotype 2 sequences in a clade (identified as * in Fig. 1; henceforth designated G2*) supported by a high bootstrap value (96%; Fig. 1A), or posterior probability (1.00; Fig. 1B). The existence of this G2* subcluster was confirmed by multivariate principal coordinate analysis (Fig. 2A). The coordinate scores obtained by PCOORD covered 19.2% of the total variation in the DNA sequences analyzed (11.8% for the first coordinate and 7.4% for the second), and indicated a clear distinction between the G2 and G2* groups of sequences. Another group of viral sequences amplified formed a separate cluster (identified as *** in Fig. 1) with three genotype 2 references. However, support for this clade was only found in the Bayesian tree (posterior probability of 1.00, Fig. 1B).

To investigate whether the G2* could be considered a new GBV-C genotype, the distributions of intra- and intergenotype pairwise genetic distances were analyzed (Fig. 2B). The dispersal of the pairwise genetic distances between G2* and all the other genotype 2 sequences produced a distribution that fell in between those obtained for the intra- and intergenotype genetic

distance comparisons (with average diversity values of 9.8% and 14.9%, respectively). Therefore, despite the fact that the topologies of the NJ and Bayesian trees were compatible with the suggestion of a new GBV-C genotype (G2*), genetic distance comparisons do not support this hypothesis. Consequently, G2* may simply correspond to a geographic subcluster within the genotype 2 radiation.

The G2* group of sequences was characterized by 7.8% of overall genetic diversity and the presence of a Q residue at position 134 in the putative E2 protein, which was not observed in any of the other GBV-C sequences analyzed (including reference sequences from all the other viral genotypes; $P < 0.001$). When compared to the other genotype 2 sequences, the presence of putative H258 and F182 was also found to be statistically associated with the G2* subcluster ($P < 0.001$). The Shannon entropy, which takes into account both the variety and frequency of observed amino acids in each position in a protein sequence, was used to describe the relative conservation in the analyzed putative protein sequences. Using G2 and G2* as background and query sequence sets, respectively, the entropy difference between both sets was calculated. The results, schematically represented in Figure 3, highlighted four sites in the G2* (A58, S110, Q134, and F182) corresponding to highly conserved residues ($P < 0.005$), while the N86 residue was significantly more conserved in G2 than in G2*.

The estimation of the ratio of non-synonymous (dN) to synonymous (dS) substitution is an important tool to study the process of sequence evolution. The dN/dS ratio was, therefore, calculated for the whole set of GBV-C sequences obtained. As previous reports in the

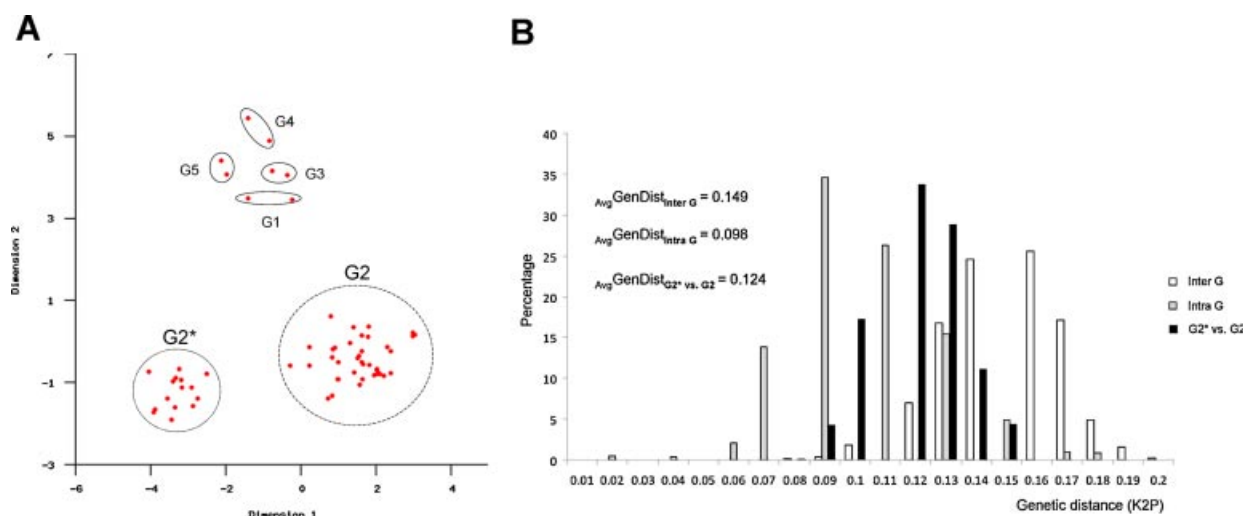


Fig. 2. (A) Multivariate principal coordinate analysis (PCOORD) for the groups G2 and G2*. Axes are the two dimensions that were first extracted; together they cover 19.2% of the total differences between the two groups. The first 10 axes cover 49.8% of variation. G1 (AY951958, U36380), G3 (AB000164, D87711), G4 (AB003292, AB018667), and G5 (AY951976, AY951956) were used as references. The nucleotide sequences in the G2 set include all those subtyped as

such in Fig. 1. The same applies to the G2* set. (B) Distribution of intragenotype (Intra G; grey bars), and intergenotype (Inter G; white bars) pairwise genetic distances, corrected with the K2P formula. The frequencies of each genetic distance interval were converted to percentages. The distribution of the pairwise genetic distance values when G2* were compared to G2 is represented by the black bars. Avg GenDist indicates average genetic distance values.

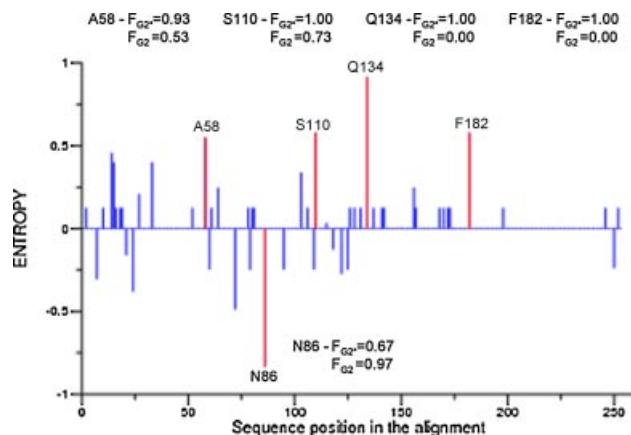


Fig. 3. Graphical representation of entropy (Ent.) differences between the G2 and G2* putative amino acid sequences. The analysed sequences correspond to a segment of the viral polyprotein covering the C-terminus of E1 and most of E2. Positive values indicated higher entropy in the background (G2) sequence set ($\text{Ent.G2} - \text{Ent.G2}^* > 0$), while negative entropy difference values indicate higher entropy in the query (G2*) sequence set ($\text{Ent.G2} - \text{Ent.G2}^* < 0$). Residues which were found to be highly conserved in G2 and G2* (P values < 0.005) are indicated along with their respective frequencies in each data set (F_{G2^*} , F_{G2}).

literature indicate [for a review see Simmonds, 2001], an overt excess of synonymous substitutions was detected, giving an estimated dN/dS ratio of 0.042. Nevertheless, in spite of a clearly steepest increase in the cumulative behavior of dS relatively to dN over the whole sequence analyzed (not shown), codon-by-codon analysis of dN/dS ratios showed that certain codons (encoding residues at positions 347 in E1, or 79, 134, 182, and 258 in E2) seemed to be under positive selection ($\text{dN/dS} \geq 2$), while in others, genetic change was highly restrained (codons 137 and 242 in E2, $\text{dN/dS} \leq 0.04$). Altogether, the putative amino acid sequences analyzed evidenced a considerable degree of conservation (average genetic distance (Poisson's correction) of 3.4%).

DISCUSSION

Despite the fact that the GBV-C virus can be regarded as relatively highly prevalent in humans, the highest rates of infection are associated with population groups at high risk of exposure to blood-borne viruses. These include injecting drug users, hemophiliacs, and hemodialysis, renal or hepatic post-transplantation patients, and those infected with HIV or HCV [reviewed by George et al., 2005; Reshetnyak et al., 2008]. Most probably as a consequence of similar routes of transmission, coinfection with HIV and/or HCV is rather common, with GBV-C viremia being detected in up to 40% of the persons coinfecting with HCV and/or HIV [George et al., 2005].

The prevalence of infection and genetic characteristics of the GBV-C virus circulating in the general Portuguese population are unknown. Therefore, to maximize the chances of detecting GBV-C, a decision was made to focus this study on a convenience sample of 214 individuals, originating from the Lisbon metropol-

itan area, and sharing high rates of infection with parenterally transmitted viruses. As expected, a high rate of GBV-C infection (40.7%) was detected by RT-PCR targeting the conserved 5'-UTR. Only 23.0% ($n = 20$) of the individuals with GBV-C viremia were monoinfected, the majority (77.0%, $n = 67$) being coinfecting with either HIV and/or HCV. GBV-C viremia was associated with the age (> 25 years) of the individuals studied, in agreement with a previous study [Reshetnyak et al., 2008]. The same holds true for the association with the use of injected drugs. However, the marginally statistical significance of the results obtained may be a consequence of the different routes contributing to infection with GBV-C in humans. Although parenteral exposure has been described as the major route of transmission for GBV-C, sexual [Stark et al., 1999; Ramia et al., 2004], vertical [Lefrère et al., 2000; Ohto et al., 2000], and communal (familial) transmission [Cheng et al., 2003] have also been described. As far as coinfection with HIV-1 is concerned, analysis of the available CD4^+ cell counts and viral load data did not indicate any protective role played by GBV-C, as suggested previously [reviewed by Shankar et al., 2008]. However, such fact may be a consequence of a small sample size and/or the variable stage of HIV-1 infection for the individuals included in this study.

Previous reports have demonstrated the existence of genogroups of GBV-C strains, frequently known as genotypes, with a distribution that correlates with geographical or ethnical origin [Muerhoff et al., 1997, 2006; Okamoto et al., 1997]. This observation has been used to support the hypothesis of GBV-C co-evolution with humans [Charrel et al., 1999]. In general terms, genotype 1 is the most frequent in Africa, genotype 2 in Europe/America. Genotypes 3–5 have been found in Japan/South America, South-East Asia, and South Africa respectively, while a putative sixth genotype was identified in Indonesia [Muerhoff et al., 2006; Reshetnyak et al., 2008]. The extent of GBV-C global diversity is, however, limited by restricted geographical sampling, as most studies have been conducted in developed countries. Consequently, little information is available for a number of regions of the world.

In this study, the genetic characterization of GBV-C sequences obtained from approximately half ($n = 44$) of the GBV-C (+) individuals ($n = 87$) was also carried out, based on the analysis of a 1.2 kb genomic fragment including 220 nucleotides (3') of E1 and most (96%) of the E2 glycoprotein coding region. The observed distribution of GBV-C Portuguese sequences between genotypes 1 and 2 (G1 and G2, respectively) was not unexpected, given the geographical position of Portugal, its long-lasting relationships with several African countries, and the high percentage of individuals reporting African nationality/ethnicity (28.5%) among the individuals studied. Indeed, detection of G1, comprehending 22.7% of the sequences analyzed, was significantly associated with African origin of the infected individuals.

The study of GBV-C genotype 2 sequences obtained from different geographical regions, as initially reported

by Lim et al. [1997], is compatible with the possibility that GBV-C dispersal occurs through multiple transmission networks, as recently described for HCV [van de Laar et al., 2009]. The results here presented are in agreement with this hypothesis. The analysis of genotype 2 sequence data evidenced clear country mixing, with approximately 56% (19/34) of the G2 sequences identified clustering in phylogenetic trees with references from Bolivia, Jamaica, USA, Japan, China, Germany, Greece, England, and Egypt. Although sampling bias may affect the geographical representation of genotype 1, its prevalence in Lisbon, most probably associated primarily with African immigration, may open new routes for the dissemination of GBV-C in Western Europe.

The G1 sequence 6388E2 segregated early from all others in the obtained phylogenetic trees (Fig. 1), suggesting a putative recombinant/divergent nature. The central part of 6388E2 was shown to correspond to an intragenotype 1 recombinant sequence, while its 5' and 3'-terminuses could not be unambiguously characterized.

Both phylogenetic methods used disclosed the segregation of a group of G2 sequences ($n = 15$) into a separate subclade (G2*), which was also supported by the results of the principal coordinate analysis (Fig. 2A). However, the distribution of genetic distances from pairwise comparisons between G2* and all other G2 sequences in the phylogenetic trees (Fig. 1) overlapped between those referring to intra- and intergenotype comparisons (Fig. 2B). Therefore, G2* seems to correspond to a geographic subcluster within the GBV-C genotype 2, rather than a new GBV-C genotype. The identification of G2* sequences in individuals living outside the Lisbon metropolitan area would be important as a means of ascertaining the geographical distribution of G2*. The identification of putative subclusters within GBV-C genotypes 1, 2, and 5 has been previously described [Muerhoff et al., 1997, 2003; Smith et al., 1997; Liu et al., 2003]. However, the lack of significant sequence overlap between the data described above and most of those used for the analysis of the G1 and G2 subclusters, precluded a definite assignment of the sequences obtained in this study to any of them.

G2* putative translation products were singled out by the presence of a Q (glutamine) residue in the protein's N-terminal half, not found in any of the other sequences analyzed. Furthermore, when compared to all the other sequences clearly assigned to genotype 2, the presence of an H (histidine) residue in the E2 C-terminal half was also found to be statistically associated with G2*. On the other hand, none of the sequences in G2* displayed the Y (tyrosine) at position 182, although this residue is frequently found in G2. Shannon entropy was calculated as a formal measure of sequence variability, and the obtained data revealed the presence of four residues (A58, S110, Q134, and F182) that were highly conserved in the G2* subcluster, while in G2, N86 was highly stable (Fig. 3). These results should be used as a guide to highlight positions which may be of biological importance.

As expected, clear conservation of the putative protein sequences analyzed was observed, as indicated by an overall average amino acid genetic distance of 3.4%, and an overt excess of synonymous over non-synonymous substitutions, reflected in a dN/dS ratio of 0.042. This suggests that the region of the GBV-C genome analyzed is under selection, and indicates that the occurrence of non-conservative changes at most codons does not seem to be favored. Without any doubt, some of these were found to be under very strong negative pressure (such as L137 and L242 in E2), but a few codons did not seem to follow this trend, and were apparently under positive selective force (encoding residues at coordinates 347 in E1, or 79, 134, 182, and 258 in E2). This suggests that despite an overall strong conservation of the protein sequence, non-conservative substitutions at a few residues may be positively selected. Since the viral glycoproteins are located at the surface of the virion, the occurrence of non-conservative changes in an overall conserved structure is most probably immune driven.

ACKNOWLEDGMENTS

This study is part of a larger research project involving the genetic characterization of blood-borne viruses circulating in the injecting drug users of the Lisbon metropolitan area. We would like to thank the staff of Centro de Atendimento de Toxicodependentes das Taipas, Av do Brasil 53, Lisboa, for their kind cooperation, and Teresa Venenno for technical assistance.

REFERENCES

- Baggio-Zappia GL, Hernandez Granato CF. 2009. HIV-GB virus C co-infection: An overview. *Clin Chem Lab Med* 47:12–19.
- Charrel RN, De Micco P, de Lamballerie X. 1999. Phylogenetic analysis of GB viruses A and C: Evidence for cospeciation between virus isolates and their primate hosts. *J Gen Virol* 80:2329–2335.
- Cheng PN, Chang TT, Jen CM, Ko AW, Young KC, Wu HL. 2003. Molecular evidence for transmission of GB virus-C/Hepatitis G virus infection within family: Close relationship between mother and child. *Hepatogastroenterology* 50:151–156.
- Esteves A, Parreira R, Venenno T, Franco M, Piedade J, Germano de Sousa J, Canas-Ferreira WF. 2002. Molecular epidemiology of HIV type 1 infection in Portugal: High prevalence of non-B subtypes. *AIDS Res Hum Retroviruses* 18:313–325.
- Esteves A, Parreira R, Piedade J, Venenno T, Franco M, Germano de Sousa J, Brum P, Costa A, Canas-Ferreira WF. 2003. Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Res Hum Retroviruses* 19:511–517.
- George SL, Varmaz D. 2005. What you need to know about GB virus C. *Curr Gastroenterol Rep* 7:54–62.
- George SL, Wünschmann S, McCoy J, Xiang J, Stapleton JT. 2002. Interactions between GB virus type C and HIV. *Curr Infect Dis Rep* 4:550–558.
- George SL, Varmaz D, Stapleton JT. 2005. GB virus C replicates in primary T and B lymphocytes. *J Infect Dis* 193:451–454.
- Hall TA. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95–98.
- Huelsenbeck JP, Ronquist F. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298.
- Kobayashi M, Tanaka E, Nakayama J, Furuwatari C, Katsuyama T, Kawasaki S, Kiyosawa K. 1999. Detection of GB virus-C/hepatitis G

- virus genome in peripheral blood mononuclear cells and liver tissue. *J Med Virol* 57:114–121.
- Leary TP, Muerhoff AS, Simons JN, Pilot-Matias TJ, Erker JC, Chalmers ML, Schlauder GG, Dawson GJ, Desai SM, Mushahwar IK. 1996. Sequence and genomic organization of GBV-C: A novel member of the flaviviridae associated with human non-A-E hepatitis. *J Med Virol* 48:60–67.
- Lefrère JJ, Sender A, Mercier B, Mariotti M, Pernot F, Soulié JC, Malvoisin A, Berry M, Gabai A, Lattes F, Galiay JC, Pawlak C, de Lachaux V, Chauveau V, Hreiche G, Larsen M, Férec C, Parnet-Mathieu F, Roudot-Thoraval F, Brossard Y. 2000. High rate of GB virus type C/HGV transmission from mother to infant: Possible implications for the prevalence of infection in blood donors. *Transfusion* 40:602–607.
- Lim MY, Fry K, Yun A, Chong S, Linnen J, Fung K, Kim JP. 1997. Sequence variation and phylogenetic analysis of envelope glycoprotein of hepatitis G virus. *J Gen Virol* 78:2771–2777.
- Liu HF, Teng CW, Fukuda Y, Nakano I, Hayashi K, Takamatsu J, Goubau P, Toyoda H. 2003. A novel subtype of GB virus C/hepatitis G virus genotype 1 detected uniquely in patients with hemophilia in Japan. *J Med Virol* 71:385–390.
- Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73:52–160.
- Muerhoff AS, Smith DB, Leary TP, Erker JC, Desai SM, Mushahwar IK. 1997. Identification of GB virus C variants by phylogenetic analysis of 5'-untranslated and coding region sequences. *J Virol* 71:6501–6508.
- Muerhoff AS, Tillmann HL, Manns MP, Dawson GJ, Desai SM. 2003. GB virus C genotype determination in GB virus-C/HIV co-infected individuals. *J Med Virol* 70:141–149.
- Muerhoff AS, Leary TP, Sathar MA, Dawson GJ, Desai SM. 2005. African origin of GB virus C determined by phylogenetic analysis of a complete genotype 5 genome from South Africa. *J Gen Virol* 86:1729–1735.
- Muerhoff AS, Dawson GJ, Desai SM. 2006. A previously unrecognized sixth genotype of GB virus C revealed by analysis of 5'-untranslated region sequences. *J Med Virol* 78:105–111.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
- Ohto H, Ujiie N, Sato A, Okamoto H, Mayumi M. 2000. Mother-to-infant transmission of GB virus type C/HGV. *Transfusion* 40:725–730.
- Okamoto H, Nakao H, Inoue T, Fukuda M, Kishimoto J, Iizuka H, Tsuda F, Miyakawa Y, Mayumi M. 1997. The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. *J Gen Virol* 78:737–745.
- Pavesi A. 2001. Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *J Mol Evol* 53:104–113.
- Radkowski M, Kubicka J, Kisiel E, Cianciara J, Nowicki M, Rakela J, Laskus T. 2000. Detection of active hepatitis C virus and hepatitis G virus/GB virus C replication in bone marrow in human subjects. *Blood* 95:3986–3989.
- Ramia S, Mokhbat J, Sibai A, Klayme S, Naman R, HIV Epidemiology Research Study Group. 2004. Exposure rates to hepatitis C and G virus infections among HIV-infected patients: Evidence of efficient transmission of HGV by the sexual route. *Int J STD AIDS* 15:463–466.
- Reshetnyak VI, Karlovich TI, Ilchenko LU. 2008. Hepatitis G virus. *World J Gastroenterol* 14:4725–4734.
- Shankar EM, Solomon SS, Vignesh R, Murugavel KG, Sundaram M, Solomon S, Balakrishnan P, Kumarasamy N. 2008. GB virus infection: A silent anti-HIV panacea within? *Trans R Soc Trop Med Hyg* 102:1176–1180.
- Simmonds P. 2001. The origin and evolution of hepatitis viruses in humans. *J Gen Virol* 82:693–712.
- Smith DB, Cuceanu N, Davidson F, Jarvis LM, Mokili JL, Hamid S, Ludlam CA, Simmonds P. 1997. Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5' non-coding region. *J Gen Virol* 78:1533–1542.
- Smith DB, Basaras M, Frost S, Haydon D, Cuceanu N, Prescott L, Kamenka C, Millband D, Sathar MA, Simmonds P. 2000. Phylogenetic analysis of GBV-C/hepatitis G virus. *J Gen Virol* 81:769–780.
- Stark K, Doering CD, Bienzle U, Pauli G, Hamouda O, Engel AM, Schreier E. 1999. Risk and clearance of GB virus C/hepatitis G virus infection in homosexual men: A longitudinal study. *J Med Virol* 59:303–306.
- Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599.
- van de Laar T, Pybus O, Bruisten S, Brown D, Nelson M, Bhagani S, Vogel M, Baumgarten A, Chaix ML, Fisher M, Gotz H, Matthews GV, Neifer S, White P, Rawlinson W, Pol S, Rockstroh J, Coutinho R, Dore GJ, Dusheiko GM, Danta M. 2009. Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. *Gastroenterology* 136:1609–1617.